

1

2 **Genome Mining Uncovers Clustered Biosynthetic Pathways for Defense-Related**

3 **Molecules in Bread Wheat**

4

5 Guy Polturak^{1*}, Martin Dippe^{1,2}, Michael J Stephenson¹, Rajesh Chandra Misra¹, Charlotte Owen¹,

6 Ricardo Ramirez-Gonzalez³, John Haidioulis³, Henk-Jan Schoonbeek³, Laetitia Chartrain³, Philippa

7 Borrill¹, David R Nelson⁴, James Brown³, Paul Nicholson³, Cristobal Uauy³ and Anne Osbourn^{1*}

8 ¹Department of Biochemistry and Metabolism, John Innes Centre, Norwich NR4 7UH, United

9 Kingdom

10 ²Current address: Department of Bioorganic Chemistry, Leibniz Institute of Plant Biochemistry, D-

11 06120 Halle (Saale), Germany

12 ³Department of Crop Genetics, John Innes Centre, Norwich NR4 7UH, United Kingdom

13 ⁴Department of Microbiology, Immunology and Biochemistry, University of Tennessee Health

14 Science Center, Memphis, TN 38163

15 *To whom correspondence should be addressed: Email: guy.polturak@jic.ac.uk;

16 anne.osbourn@jic.ac.uk

17

18

19 **ABSTRACT**

20 Wheat is one of the most widely grown food crops in the world. However, it succumbs to numerous
21 pests and pathogens that cause substantial yield losses. A better understanding of biotic stress responses
22 in wheat is thus of major importance. Here we identify previously unknown pathogen-induced
23 biosynthetic pathways that produce a diverse set of molecules, including flavonoids, diterpenes and
24 triterpenes. These pathways are encoded by six biosynthetic gene clusters and share a common
25 regulatory network. We further identify associations with known or novel phytoalexin clusters in other
26 cereals and grasses. Our results significantly advance the understanding of chemical defenses in wheat
27 and open up new avenues for enhancing disease resistance in this agriculturally important crop.

28

29 **INTRODUCTION**

30 The allohexaploid bread wheat (*Triticum aestivum*) accounts for approximately 20% of the calories
31 consumed by humans worldwide¹. Around one fifth of the global annual wheat yield is lost due to pest
32 and pathogen attack², a value that is expected to sharply rise as the climate warms^{3,4}. A better
33 understanding of how wheat responds to biotic stresses could enable the development of strategies for
34 minimizing yield losses and reducing reliance on pesticides. Significant advances have been made in
35 identification of wheat resistance genes (R-genes) involved in pathogen recognition and the immune
36 response⁵. However, very little is known about the chemical defenses produced by wheat in response
37 to pathogen attack (phytoalexins).

38 The agronomic importance of wheat has led to extensive research into its genetics, and to the
39 generation of a vast body of transcriptome data from numerous studies into wheat development,
40 physiology, and interactions with the environment. However, the first bread wheat genome assembly
41 became available only recently because of the challenges associated with its large genome size, high
42 repetitive sequence content, and relatedness between homoeologous subgenomes⁶. The availability of
43 these genome and transcriptomic resources now offers the opportunity to employ a genomics-driven
44 approach to uncover novel chemical defense molecules and biosynthetic pathways in this valuable crop.

45 Such an approach is particularly useful for uncovering metabolites that are produced in small quantities
46 or under specific conditions (e.g. pathogen-induced), thereby eluding traditional chemical analyses⁷.

47 Here, by coupling gene co-expression network analysis with genome mining, we identify six
48 defense-related candidate biosynthetic gene clusters (BGCs) in bread wheat. We show by expression of
49 cluster genes in *Nicotiana benthamiana* that these BGCs encode pathways for the production of
50 flavonoid, diterpene and triterpene compounds that likely serve as broad-spectrum phytoalexins in
51 wheat. Through comparative genomics we also identify associations with known or novel phytoalexin
52 clusters in other cereals and grasses. We further report the full characterization of the pathways for the
53 novel defence compounds ellarinacin and brachynacin, which are respectively produced by related gene
54 clusters in wheat and the grass purple false brome (*Brachypodium distachyon*). Our work uncovers new
55 biosynthetic pathways for novel pathogen-induced compounds in wheat and demonstrates a powerful
56 approach for rapid discovery of defense-related molecules and metabolic pathways in crop plants, which
57 may have future applications in crop protection.

58

59 **RESULTS**

60 **Gene co-expression network analysis coupled with genome analysis identifies candidate** 61 **pathogen-induced biosynthetic gene clusters in wheat**

62 In a recently published study, 850 transcriptome datasets were compiled and analyzed to produce a
63 genome-wide view of homoeolog expression patterns in hexaploid bread wheat. Weighted Gene Co-
64 expression Network Analysis (WGCNA) was carried out based on gene expression patterns in the
65 compiled datasets, and an additional set of networks were built for six separate subsample sets: grain,
66 leaf, spike, root, abiotic and disease⁸.

67 We hypothesized that new defense-related metabolites and metabolic pathways in wheat could
68 be found by mining the ‘disease’ gene network. Specifically, genes that are physically clustered in the
69 genome and are co-induced by pathogens or pathogen-associated molecules could serve as excellent
70 candidates for biosynthesis of defense compounds⁷. WGCNA assigned 55,646 genes from the ‘disease’
71 network (generated from 163 RNA-seq samples) into 69 modules based on their expression, and

72 expression values of all genes in each module were averaged to get a single ‘eigengene’ expression
73 pattern per module⁸. To find genes that exhibit a general, non-specific induction by exposure to
74 pathogens or pathogen-associated molecular patterns (PAMPs), we averaged for each module the
75 difference in normalized eigengene expression between treatment and control in seven different studies
76 and sorted the modules by the average expression delta (Fig. 1a). The top five modules (i.e., the modules
77 represented by the most highly induced eigengenes), namely ME34, ME25, ME12, ME36 and ME8,
78 showed consistent induction in all seven experiments used in the analysis (Fig. 1b) and were selected
79 for further investigation.

80 To determine whether any of these gene expression modules contained genes that form putative
81 biosynthetic gene clusters, we next mined the five modules by filtering for groups of three or more
82 genes with successive accession numbers, i.e. that are physically adjacent in the genome. A total of 55
83 groups were found (Extended Data Table 1), which include groups of tandem duplicates as expected.
84 Twenty contain protein kinase genes with possible roles in biotic stress responses. A further six consist
85 of genes for different types of enzyme families associated with plant specialized metabolism, and so
86 were identified as possible biosynthetic gene clusters (BGCs) for synthesis of defense compounds.
87 These six putative BGCs included two pairs of homoeologous clusters and were thus defined as four
88 cluster types (cluster types 1-4), and assigned as clusters 1(2A), 1(2D), 2(2B), 3(5A), 3(5D), and 4(5D).
89 The bracketed numbers refer to the chromosomes that the clusters are located on (Fig. 1c,
90 Supplementary Table 1). The two homoeologous cluster pairs are 1(2A) and 1(2D), and 3(5A) and
91 3(5D), respectively.

92 The majority of genes in the six putative BGCs were found in a single module, ME25,
93 indicating highly similar expression patterns and suggesting possible co-regulation of the BGCs by a
94 shared network of transcription factors (TFs) (Supplementary Table 1). Analysis of a previously
95 generated GENIE3-based wheat regulatory network⁸ indeed revealed a highly overlapping network of
96 TFs predicted to interact with the six BGCs. Specifically, 137 TFs predicted to interact with genes from
97 two or more of the BGCs were found, including 21 TFs from 10 groups (i.e. groups of homoeologs or
98 tandem duplicates) predicted to interact with genes from all six clusters. The top five most highly
99 interacting TF groups included transcription factors from the WRKY, bHLH (two groups), NAC, and

100 HSF families (Fig. 1d, Extended Data Table 2), all of which have been associated with regulation of
101 phytoalexin biosynthesis or pathogen resistance in plants^{9,10}. Examination of Gene Ontology (GO) term
102 enrichment of the predicted target genes of representative TFs from each of the five groups showed that
103 the most significantly enriched terms are related to immune response or defense from biotic stress, for
104 all five TFs excluding the NAC transcription factor, for which the most significantly enriched GO terms
105 were related to response to chemicals/toxins (Extended Data Table 2). Of the 21 TFs that are associated
106 with all six BGCs, none interact with any of the characterized genes for the biosynthetic pathway of the
107 benzoxazinoids¹¹ (e.g., DIBOA, DIMBOA), a group of well-characterized defense compounds found
108 in several cereal crops, including wheat¹² (Extended Data Table 2). This is consistent with the definition
109 of benzoxazinoids as phytoanticipins (constitutively produced defense compounds)¹³, also reflected by
110 the fact that the benzoxazinoid biosynthetic genes are not found in the pathogen-induced WGCNA
111 expression modules.

112

113 **The six predicted biosynthetic gene clusters comprise co-expressed genes potentially involved in**
114 **diterpene, triterpene and flavonoid metabolism**

115 Plant BGCs typically contain one or more genes required for generation of a natural product scaffold,
116 along with genes encoding downstream tailoring enzymes that modify this scaffold (e.g. cytochromes
117 P450 (CYPs), sugar transferases (UGTs), methyl transferases (MTs))¹⁴. The six predicted pathogen-
118 induced wheat BGCs each contain 5-7 co-expressed biosynthetic genes (Fig.1c, Supplementary Fig. 1).
119 Based on the gene annotations, the predicted scaffold-forming enzymes for the clusters are terpene
120 synthases (TPSs) (clusters 1(2A), 1(2D) and 2(2B)); oxidosqualene cyclases (OSCs) (clusters 3(5A)
121 and 3(5D)); and a chalcone synthase (CHS) (cluster 4(5D)), hallmarks of diterpene, triterpene and
122 flavonoid biosynthesis, respectively. Notably, all three classes of compound are associated with plant
123 defense, including in the grasses¹⁵⁻¹⁸.

124 Co-expression within each cluster was assessed by calculation of the Pearson correlation
125 coefficient (r-val) between the expression of a representative scaffold-forming gene from each cluster
126 and other cluster genes, within an RNA-seq dataset including 68 experiments from the wheat-

127 expression.com website^{8,19}. In the putative diterpene clusters 1(2A) and 1(2D), several genes were found
128 to be highly co-expressed with the TPS bait ($r\text{-val}>0.8$), including a copalyl diphosphate synthase
129 (CPS), encoding a key enzyme in diterpene biosynthesis that typically catalyzes the preceding step to
130 TPS; one 1(2D) or two 1(2A) UGTs; and three CYPs. In cluster 2(2B), two TPSs, two CYPs and a CPS
131 are co-expressed. In clusters 3(5A) and 3(5D) all five genes are co-expressed, while in cluster 4(5D) all
132 genes are co-expressed with the exception of one chalcone synthase duplicate and a chalcone-flavanone
133 isomerase (Supplementary Table 1, Fig. 1c).

134

135 **The type 4 biosynthetic gene cluster 4(5D) encodes a functional flavonoid biosynthetic pathway**

136 To establish whether the predicted BGCs were likely to be functional, we first investigated the candidate
137 flavonoid BGC 4(5D) (Fig. 1c, Supplementary Table 1, Supplementary Fig. 2). The genes for the
138 predicted scaffold-generating enzyme (TaCHS1) and co-expressed tailoring enzymes (TaCYP71C164
139 and TaOMT3/6/8) were cloned and transiently expressed in *Nicotiana benthamiana* by
140 agroinfiltration²⁰, together with the clustered chalcone-flavanone isomerase (chi-D1), and an additional
141 CYP71 gene (TaCYP71F53_5D), which is located 425 Kb upstream of the terminal OMT of the cluster
142 and also belongs to the ME25 expression module. A fourth OMT in the cluster (TaOMT7) is a tandem
143 duplicate of TaOMT6 with a single amino acid difference and was not included in the analysis. The
144 combined expression of all genes resulted in formation of a new product exhibiting UV absorbance
145 ($\lambda_{\text{max}} = 260\text{nm}$) with exact mass $[M+H]=329.1010$ (Supplementary Fig. 2) and predicted elemental
146 composition $\text{C}_{18}\text{H}_{17}\text{O}_6$ (-1.78 ppm). This product was not produced in combinations in which TaCHS1
147 or any of the two CYPs and three OMTs were omitted, indicating that the proteins encoded by all six
148 genes are enzymatically active (Fig. 1e, Supplementary Fig. 2). Inclusion of chi-D1 was not essential
149 for formation of this product in *N. benthamiana* (Supplementary Fig. 2). Thus, the co-expressed genes
150 within cluster 4(5D) encode a functional pathway that, based on UV absorbance, exact mass and the
151 calculated elemental composition of the putative end-product, is likely to produce a hydroxy-
152 trimethoxy-flavone. Future work is needed to fully elucidate the structures of the pathway end-product
153 and intermediates.

154

155 **The homoeologous type 1 biosynthetic gene clusters 1(2A) and 1(2D) are related to but**
156 **functionally distinct from the rice momilactone cluster**

157 Rice produces a variety of diterpene phytoalexins for which the biosynthetic pathways are well-
158 characterized. The genes for several of the pathways for labdane-related diterpenes (e.g. momilactones,
159 phytocassanes/oryzalides) are clustered in the rice genome²¹⁻²³. These labdane-related diterpenes are
160 formed from the universal diterpenoid precursor geranylgeranyl diphosphate (GGPP) (**1**) via initial
161 cyclization reactions catalyzed by copalyl diphosphate synthases (CPSs) that produce normal, *ent*, or
162 *syn* stereoisomers of copalyl diphosphate (CPP). The CPP intermediates are subsequently utilized by
163 terpene synthases (TPSs) to form various diterpene backbones, which then typically undergo further
164 tailoring reactions²⁴. In wheat, diterpene metabolism is considerably less well characterized than in rice.
165 In previous studies aimed at functional characterization of diterpene-related genes in wheat, five copalyl
166 diphosphate synthases (CPS1-5) and six kaurene synthase-like terpene synthases (KSL1-6) were cloned
167 and characterized by recombinant expression²⁵⁻²⁷. Four of the CPS enzymes catalyzed production of
168 normal or *ent* stereoisomers of CPP, while five of the KSL enzymes were shown to convert normal-,
169 *ent*-, or *syn*- CPP to several different diterpene products. The physical location and general expression
170 patterns of these genes were, however, unknown.

171 Interestingly, our study identified two of these genes, namely *TaKSL1* and *TaCPS2*, as the co-
172 expressed TPS and CPS genes in cluster 1(2A). A third gene, *TaKSL4* is found in the homoeologous
173 1(2D) cluster (Fig. 2a). *TaKSL4* is not co-expressed with other cluster genes and generally exhibits a
174 root-specific, non-induced expression pattern (Supplementary Fig. 1). The co-localization and co-
175 expression of *TaKSL1* and *TaCPS2* coincides with their previously ascribed enzymatic functions -
176 *TaCPS2* produces normal-CPP (**2**), while *TaKSL1* acts on normal-CPP to produce isopimara-7,15-
177 diene (**3**)^{25,26}. *TaKSL1* can also react with a *syn*-CPP substrate, but a *syn*-CPP producing copalyl
178 synthase is yet to be identified in wheat²⁶. Transient expression of the Chr.2D homoeologs of *TaCPS2*
179 and *TaKSL1* (named *TaCPS-D2* and *TaKSL-D1* hereinafter) in *N. benthamiana* revealed that these
180 enzymes are functional and produce compounds with mass spectra matching copalol and isopimara-

181 7,15-diene respectively, confirming the activity of this pair of genes in the 1(2D) cluster (Fig 2b,
182 Supplementary Fig. 3). The occurrence of additional co-expressed CYP genes and a UGT gene in the
183 1(2D) and 1(2A) clusters (Fig. 2a) suggests that these clusters form pathogen-induced pathways for
184 production of isopimara-7,15-diene-derived diterpenes (Fig 2c).

185 Intriguingly, microsynteny analysis between wheat and rice suggests that the type 1 clusters
186 present on wheat chromosomes 2A and 2D (BGCs 1(2A) and 1(2D)) likely share a common
187 evolutionary origin with the rice momilactone cluster. The *KSL* genes in clusters 1(2A) and 1(2D) are
188 close homologs of the *OsKS4* gene from the rice momilactone BGC^{21,22}. Directly adjacent to *TaKSL1*
189 is a *CPS* gene that is orthologous to *OsCPS4*. The wheat cluster also includes four cytochrome P450s
190 belonging to the CYP99 family that are homologs of the CYP99A2/A3 P450 pair in the rice
191 momilactone BGC²². Furthermore, the chromosomal regions harbouring wheat clusters 1(2a) and 1(2D)
192 are syntenic to the region of the rice genome containing the momilactone cluster, which is found on rice
193 Chr.4, the corresponding chromosome of wheat Chr.2²⁸ (Fig. 2d). However, although these clusters may
194 share a common evolutionary origin, they produce different types of diterpenes: the rice momilactones
195 are derivatives of the *syn*-CPP-derived scaffold *syn*-pimara-7,15-diene²¹, while functional
196 characterization and gene expression data of the wheat 1(2D) cluster and previous characterization of
197 the *TaKSL1* and *TaCPS2* genes^{25,26}, which we have shown to be in wheat cluster 1(2A), implies that
198 these two BGCs encode pathways that yield derivatives of the normal-CPP-derived isopimara-7,15-
199 diene-scaffold. Of note, the rice momilactone cluster also includes two short-chain
200 dehydrogenase/reductase (SDR) genes *OsMAS* and *OsMAS2*^{22,29} that do not have apparent orthologs in
201 the wheat type 1 clusters or elsewhere in the wheat genome.

202 The third predicted diterpene BGC that we found, cluster 2(2B), also includes three other
203 previously characterized wheat genes, namely *TaCPS1*, *TaKSL2* and *TaKSL3*^{25,26}, all of which are co-
204 expressed (Fig. 2e). *TaCPS1* catalyzes formation of *ent*-CPP (**4**), while *TaKSL2* acts on *ent*-CPP to
205 produce pimara-8(14),15-diene (**5**). *TaKSL3*, a tandem duplicate of *TaKSL2*, only exhibits low activity,
206 selectively acting on *ent*-CPP to produce two unknown products²⁶. The combined functions of *TaCPS1*
207 and *TaKSL2*, together with the presence of additional co-expressed CYPs in the cluster, suggest that
208 BGC 2(2B) encodes a pathway for production of *ent*-pimara-8(14),15-diene derivatives (Fig. 2c).

209

210 **The homoeologous type 3 cluster 3(5D) encodes a biosynthetic pathway to ellarinacin, a novel**
211 **arborinane-type triterpenoid**

212 The type 3 cluster 3(5D) contains genes implicated in triterpenoid biosynthesis, most notably a
213 predicted oxidosqualene cyclase gene (*TaOSC*). Flanking *TaOSC* are three cytochrome P450s
214 (*TaCYP51H35*, *TaCYP51H37* and *TaCYP51H13P*) and a gene annotated as a 3 β -hydroxysteroid-
215 dehydrogenase/decarboxylase (*TaHSD*) (Fig. 3a). The genomic sequences of *TaOSC*, *TaHSD*,
216 *TaCYP51H35* and *TaCYP51H37* predict full coding sequences for all four genes, while *TaCYP51H13P*
217 was found by manual annotation to carry two premature stop codons (Supplementary Fig. 4) and was
218 designated a pseudogene. The homoeologous cluster on Chr.5A is similarly structured (Fig. 3a), but
219 with a predicted full coding sequence for *TaCYP51H13_5A*. Amino acid sequence identity between
220 homoeologous pairs in the 3(5D) and 3(5A) clusters is >99% for *TaOSC*, *TaHSD* and *TaCYP51H37*,
221 and >97% for *TaCYP51H35* and *TaCYP51H13*. As for the type 1 diterpene cluster, homoeologs of the
222 type 3 cluster genes are not found in the B genome. However, a homologous gene cluster is present
223 adjacent to the 3(5D) cluster on Chr.5D, which includes paralogs of the *TaOSC*, *TaHSD*, and
224 *TaCYP51H* genes. Similarly, one *TaOSC* and two *CYP51H* paralogs are also found on Chr.5A, adjacent
225 to the 3(5A) cluster (Supplementary Fig. 5). These Chr.5A and Chr.5D paralogs, however, in general
226 have low expression across all transcriptomic data available on wheat-expression.com, and so are not
227 likely to belong to active BGCs (Supplementary Table 2).

228 Functional analysis of the cluster 3(5D) genes was carried out by transient expression in *N.*
229 *benthamiana*. All genes were co-infiltrated with an *Agrobacterium* strain harbouring an expression
230 construct for a feedback insensitive form of 3-hydroxy-3-methylglutaryl coenzyme A reductase
231 (tHMGR) from oat, which enhances triterpenoid precursor supply³⁰. GC- and LC-MS analyses of leaf
232 extracts revealed that the four enzymes *TaOSC*, *TaHSD*, *TaCYP51H35* and *TaCYP51H37* form a
233 sequential biosynthetic pathway (Fig. 3b, Supplementary Figs. 6-9). As the *TaCYP51H13P* pseudogene
234 from cluster 3(5D) does not encode a complete functional protein, we tested the activity of its Chr.5A
235 homoeolog, *TaCYP51H13_5A*, through agroinfiltration with the four 3(5D) cluster genes in different

236 combinations. TaCYP51H13_5A exhibited the same activity as TaCYP51H35, but to a lower extent,
237 resulting in lower levels of product compared to TaCYP51H35 (Supplementary Fig. 10). This redundant
238 activity provides a possible explanation why *TaCYP51H13* is not conserved in the Chr.5D cluster.

239 The structures of the purified products of co-expression of TaOSC+TaCYP51H35, and of the
240 combined four cluster genes (all from cluster 3(5D)) were determined by NMR following large-scale
241 vacuum-mediated agroinfiltration and purification (Supplementary Figs. 11,12, Supplementary Tables
242 3, 4). The product of co-expression of TaOSC+TaCYP51H35 was identified as 19-hydroxy-
243 isoarborinol (**8**), indicating that TaOSC (hereinafter isoarborinol synthase, TaIAS) synthesizes the
244 triterpene scaffold isoarborinol (**7**) which is subsequently hydroxylated by TaCYP51H35 (hereinafter
245 isoarborinol 19-hydroxylase, TaIAH). The product of co-expression of all four cluster genes was found
246 to have an unusual triterpenoid structure, with a β -epoxy group and an ether bridge attached to the A
247 ring (Fig. 3c). The GC/LC-MS data and NMR-assigned structure together suggest oxidation of the 3-
248 alcohol to the ketone 19-hydroxy-isoarborinone (**9**) by TaHSD (hereinafter 19-hydroxy-isoarborinol
249 dehydrogenase, TaHID); TaCYP51H37 (hereinafter 19-hydroxy-isoarborinone oxidase, TaHIO) likely
250 then hydroxylates the C25-methyl carbon, leading to nucleophilic attack of the A-ring ketone, thus
251 forming a hemiacetal intermediate which further reacts to produce the epoxide. This unusual reaction
252 may involve two independent catalytical cycles mediated by TaHIO. This would be in line with the
253 only other previously reported non-canonical CYP51 enzyme (AsCYP51H10, *Sad2*) which
254 hydroxylates the C16 position of the β -amyrin scaffold and also converts an alkene to an epoxide at
255 C12-C13 via two independent reactions³¹. However, a mechanism involving just one catalytic cycle
256 may also be possible (Supplementary Fig. 13). The structure of the BGC 3(5D) product has not, to the
257 best of our knowledge, been previously reported, and was named ellarinacin (**10**). The proposed
258 biosynthetic pathway is shown in Fig. 3c.

259 Interestingly, the ellarinacin cluster (BGC 3(5D)) provides multiple links to sterol metabolism.
260 Production of plant sterols from 2,3-oxidosqualene (**6**) is initiated by highly conserved OSCs known as
261 cycloartenol synthases (CASs), while triterpene scaffolds are generated from 2,3-oxidosqualene by
262 other diverse OSCs (triterpene synthases)¹⁶. TaOSC shares higher sequence similarity with
263 characterized monocot CAS enzymes in comparison to other functionally characterized monocot

264 triterpene synthases (Supplementary Figs. 14, 15). Plant 3 β -hydroxysteroid-
265 dehydrogenase/decarboxylases belong to the short chain dehydrogenase reductase (SDR) superfamily
266 and are involved in biosynthesis of phytosterols and steroidal glycoalkaloids³²⁻³⁵. Phylogenetic analysis
267 shows that TaHSD is related to *Arabidopsis thaliana* genes *3 β HSD/D1* and *3 β HSD/D2*, that take part
268 in sterol biosynthesis (Supplementary Fig. 16)^{32,36}. The cytochrome P450 genes found in the cluster
269 provide further connections to sterol metabolism, as they belong to the sterol-related CYP51 family
270 (Supplementary Figs. 17,18). CYP51 enzymes catalyze 14 α -demethylation of sterols in all eukaryotes
271 and are the only family of cytochrome P450s that are evolutionarily conserved from prokaryotes
272 through fungi, plants, and mammals³⁷. To date, only one plant CYP51 has been found to catalyze a
273 reaction different from the canonical sterol demethylase activity- AsCYP51H10 (*Sad2*), which is
274 involved in biosynthesis of an antifungal triterpene glycoside known as avenacin in oat^{31,38}. Several
275 members of the ellarinacin cluster thus appear to have been recruited from sterol biosynthetic genes,
276 most likely through gene duplication and neofunctionalization.

277

278 **The ellarinacin cluster is highly induced by biotic stress**

279 We next sought to determine whether the type 3 clusters are likely to be involved in plant defense by
280 further investigating the expression patterns of the clustered genes. Analysis of the wheat-
281 expression.com dataset revealed that the expression patterns of the genes were consistent with their
282 positioning in the ME25 and ME34 modules, i.e. that they showed induction by various fungal
283 pathogens and by the PAMPs chitin and flg22. (Supplementary Fig. 19). Notably, the clusters were not
284 substantially induced in response to various abiotic stresses, including drought, heat, cold, phosphate
285 starvation and drought-simulating treatment with PEG-6000 (Supplementary Fig. 20).

286 The observed expression pattern of the type 3 clusters was further supported by
287 semiquantitative RT-PCR analysis of selected genes from BGC 3(5A) and 3(5D) using homoeolog-
288 specific primers: expression of all tested genes was strongly induced in leaves infected with powdery
289 mildew (Li) but not by mechanical wounding (Lw), with little or no expression in the other various
290 wheat tissues analyzed (Fig. 3d). Weak induction was also observed in roots infected with

291 *Gaeumannomyces graminis*, a soil-borne fungus that causes ‘take-all’ disease (Ti). Induction of the
292 entire BGC 3(5D) by infection with powdery mildew was further validated by quantitative real-time
293 PCR (qRT-PCR). Detached wheat leaves were exposed to spores of either wheat-adapted (*Blumeria*
294 *graminis* f. sp. *tritici*, Bgt) or non-adapted (*Blumeria graminis* f. sp. *hordei*, Bgh) isolates of powdery
295 mildew, and relative transcript abundance was determined 12 and 24 hours post infection. Treatment
296 with Bgt or Bgh resulted, in both cases, in strong induction of the four cluster genes. Interestingly,
297 induction was more marked for Bgh (non-adapted) compared to Bgt (Fig. 4a). Our analyses of
298 transcriptome data from previously published studies^{39,40} in which wheat plants were challenged with
299 the fungal pathogens powdery mildew, cereal blast (*Magnaporthe* spp.), and leaf or yellow rust
300 (*Puccinia* spp.) also revealed stronger induction of the cluster genes by non-host vs. host interactions
301 (Supplementary Figs. 21, 22).

302 Finally, we analyzed gene expression in detached wheat leaves treated with the elicitors methyl
303 jasmonate (MeJa) and salicylic acid (SA), as well with the PAMP, chitin. All four cluster 3(5D) genes
304 analyzed (*TaOSC*, *TaHSD*, *TaCYP51H35*, *TaCYP51H37*) were significantly induced compared to the
305 control 12 hrs after treatment with MeJa, SA or chitin, with the exception of *TaCYP51H35* in SA-
306 treated leaves (Fig. 4b). Thus, the ellarinacin cluster is highly induced by biotic stress, suggesting a
307 possible function in wheat response against pathogens. The very low basal expression in various wheat
308 tissues, as observed in the RT-PCR and RNA-seq data analysis, and strong induction by pathogens,
309 defense-related hormones and PAMPs, further suggests that ellarinacin serves as a phytoalexin rather
310 than a phytoanticipin. Correspondingly, GC-MS analysis detected ellarinacin in extracts of MeJa-
311 treated but not control detached wheat leaves (Fig. 4c). A 60% increase in isoarborinol levels was also
312 observed in MeJa-treated leaves compared to control leaves (Supplementary Fig. 23).

313

314 **The ellarinacin cluster is conserved in wheat ancestors**

315 We next investigated whether ellarinacin-like clusters also exist in the genomes of ancestral species of
316 common wheat. Specifically, we looked for related clusters in two wild progenitors that have sequenced
317 genomes; *Aegilops tauschii* (Tausch’s goatgrass; donor of the D genome of bread wheat), and *Triticum*

318 *turgidum* ssp. *dicccoides* (wild emmer wheat, progenitor of cultivated emmer; the donor of the A and
319 B genomes of bread wheat). Microsynteny analysis of the regions surrounding the homoeologous type
320 3 BGCs on Chr.5A and Chr.5D show that while these clusters appear to be conserved on chromosome
321 5 of the A and D genomes of *A. tauschii* and wild emmer wheat, a homoeologous cluster could not be
322 found on chromosome 5B of bread wheat or wild emmer wheat. Chromosome 5B of both species do,
323 however, contain homoeologs of the OSC and/or P450s of the paralogous, transcriptionally non-active
324 cluster in the A and D genomes (Fig. 4d). The wild emmer wheat and *A. tauschii* clusters each contain
325 an OSC, an HSD and three CYP51 genes, in the same order and orientation as in wheat (Fig. 5a).
326 Sequence comparison of the cluster genes in wheat and its two wild progenitors revealed that the
327 predicted protein sequences are also highly conserved (>99.4% amino acid identity for all proteins in
328 both species; Supplementary Table 5). To assess the functionality of the *A. tauschii* cluster, we
329 transiently expressed the first two genes of the predicted *A. tauschii* pathway, namely the orthologs of
330 *TaIAS* and *TaIAH*, in *N. benthamiana*. Co-expression of the two genes resulted in formation of 19-
331 hydroxy-isoarborinol (Supplementary Fig. 24), the same product obtained by *TaIAS* and *TaIAH*
332 expression. The coding sequence of the *A. tauschii* ortholog of *TaCYP51H13P* contains one of the
333 premature stop codons found in its wheat homolog (Supplementary Fig. 4), and so is likely to be non-
334 functional. The remaining predicted active enzymes in the *A. tauschii* pathway, orthologs of TaHID and
335 TaHIO, exhibit 100% amino acid identity with their wheat counterparts, suggesting that the *A. tauschii*
336 cluster encodes a complete biosynthetic pathway for ellarinacin.

337

338 **Arborinane-type clusters are found in other grasses**

339 The occurrence of conserved ellarinacin-like clusters in wheat and its progenitors raised the possibility
340 that BGCs for ellarinacin or other arborinane-type terpenoids may also occur in other grasses. The
341 isoarborinol scaffold has been reported from other Poaceae species, including sorghum⁴¹ and rice⁴². We
342 therefore searched for orthologs of *TaIAS* in additional Poaceae species, based on sequence similarity.
343 Orthologs for *TaIAS* could not be identified in maize, sorghum, barley, and rice. The latter has a

344 previously characterized isoarborinol synthase gene⁴², but this gene bears low similarity to *TaIAS* (56%
345 similarity on amino acid level) and has most likely evolved independently.

346 A BlastP search of *TaIAS* against the recently published genome of the diploid oat species
347 *Avena strigosa*⁴³ found a candidate OSC gene on chromosome 1, herein named *AsOSC1*, with high
348 predicted amino acid sequence similarity to the *TaIAS* protein (91.2%). This was also the reciprocal
349 best hit (RBH) of *TaIAS*. Flanking *AsOSC1* (~25 Kb away) is a *CYP51H* gene, herein named
350 *CYP51H73*. Transient expression of *AsOSC1* in *N. benthamiana* yielded a new product, which was
351 verified by GC-MS as isoarborinol. No additional products were detected when *AsCYP51H73* was co-
352 expressed with *AsOSC1* (Supplementary Fig. 25). Since *AsCYP51H73* is orthologous with wheat
353 *TaCYP51H37* (*TaHIO*), we also tested if *AsCYP51H73* would exhibit the same or similar activity, by
354 co-expressing *AsCYP51H73* together with *TaIAS*, *TaIAH* and *TaHID*. However, no activity was
355 detected.

356 In the genome of the grass model plant *Brachypodium distachyon* (strain *Bd21*)⁴⁴, a *TaIAS*
357 homolog was identified on chromosome 3, *BdOSC2*, which was the RBH of *TaIAS*. Flanking this gene
358 were genes predicted to encode another highly similar OSC (*BdOSC1*), and three cytochrome P450s of
359 the CYP51H subfamily (*BdCYP51H14*, *BdCYP51H15* and *BdCYP51H16*) (Fig. 5a). A predicted
360 BAHD-type acyltransferase gene (*BdACT*) was also found between *BdCYP51H14* and *BdCYP51H15*.
361 Thus, together these genes form a potential BGC for production of arborinane-type or similar
362 triterpenoids in *B. distachyon*. A conserved cluster that has a similar gene structure to the *B. distachyon*
363 BGC but with one OSC gene only was also found in the genome of the closely related species *B. stacei*
364 (Fig. 5a).

365

366 **The *B. distachyon* Chr.3 BGC is induced by fungal pathogens**

367 To test whether the clustered genes identified on chromosome 3 of *B. distachyon* might form an active
368 BGC, their expression profiles were examined. Analysis of *B. distachyon* gene expression datasets in
369 JGI Gene Atlas (<https://phytozome.jgi.doe.gov/>),⁴⁵ and PlaNET ([http://aranet.mpimp-
370 golm.mpg.de/](http://aranet.mpimp-golm.mpg.de/))^{46,47}, showed that the three *CYP51* and two *OSC* genes are co-expressed, with highest

371 expression in the mature leaf and stem base. The BAHD acyltransferase gene displayed a similar
372 pattern, but with markedly lower overall expression values (Supplementary Fig. 26). Relative
373 expression of all six genes was further assessed by semi-quantitative RT-PCR of seven *B. distachyon*
374 tissues at different developmental stages. The cluster genes generally exhibited highest expression in
375 the leaves and stem base (Fig. 5b). A *BdACT* amplicon could only be detected with extended exposure
376 (Supplementary Fig. 27). Unlike the other cluster genes, *BdOSCI* was also expressed in the spikes of
377 mature plants.

378 qRT-PCR analysis of *B. distachyon* plants infected with *Fusarium graminearum* causing
379 Fusarium head blight or Fusarium root rot showed that, as for the wheat cluster, the *B. distachyon* cluster
380 is highly induced by fungal pathogens. Significant increases in gene expression following infection
381 were observed in both experiments for all clustered genes except *BdOSCI* (Fig. 5c).

382

383 **The *B. distachyon* Chr.3 BGC produces an arborinane-type triterpenoid**

384 Since gene expression analysis suggested an active BGC, we next investigated the functions of the
385 cluster genes by transient expression in *N. benthamiana* (Fig. 6a). GC-MS analysis revealed that
386 BdOSC2 and CYP51H15 exhibit the same activities as their respective wheat orthologs, i.e., production
387 of isoarborinol and its 19-hydroxylated derivative (Supplementary Fig. 28). Co-expression of BdOSC2
388 and CYP51H15 together with the two additional CYP51s and the BdACT acyltransferase resulted in
389 formation of the putative BGC end product, with a mass signal of [M+H-H₂O=515.3] (Supplementary
390 Figs. 29, 30). This product was purified following large-scale transient expression of the *B. distachyon*
391 cluster genes in *N. benthamiana* and found by ¹H and ¹³C NMR analyses to be an isoarborinol-derived
392 triterpenoid with hydroxyl groups on the C7,19,28 carbons and an acetoxy group on the C1 carbon.
393 (Supplementary Fig. 31, Supplementary Table 6). The assigned structure allowed the full elucidation
394 of the biosynthetic pathway from 2,3 oxidosqualene, in which BdOSC2 and BdCYP51H15 generate
395 19-OH-isoarborinol, BdCYP51H14 hydroxylates the C7 and C28 carbons to give 7,19,28-trihydroxy-
396 isoarborinol (**11**), and BdCYP51H16 hydroxylates the C1 carbon to give 1,7,19,28-tetrahydroxy-
397 isoarborinol (**12**), which is further acetylated by BdACT (Fig. 6b). This compound, which has not

398 previously been reported, was named brachynacin (**13**). The occurrence of brachynacin in *B. distachyon*
399 was verified by GC-MS analysis of leaf extracts. As for ellarinacin in wheat, the relative abundance of
400 brachynacin, as well as of isoarborinol, were found to be significantly higher in MeJa-treated vs. non-
401 treated detached leaves (Fig. 6c, Supplementary Fig. 32).

402

403 **Combination of ellarinacin and brachynacin biosynthetic genes yields novel compounds**

404 The similarities between the ellarinacin and brachynacin BGCs indicates that they possibly originate
405 from a common ancestral cluster but have evolved to produce different end products, through
406 duplication and neofunctionalization of CYP51H enzymes and recruitment of additional modifying
407 enzymes (TaHSD and BdACT, respectively). The evolution of these two pathways may have been
408 facilitated by a degree of promiscuity that enabled the pathway enzymes to accept different substrates.
409 Indeed, co-expression of different combinations of genes from the two BGCs in *N. benthamiana* did
410 yield new products. Expression of TaHSD with BdCYP51H14 and BdCYP51H15 led to production of
411 a new compound with a molecular mass $[M+H-H_2O=455.3]$, matching the expected product, 7,19,28-
412 trihydroxy-isoarborinone (Fig. 6d). Likewise, expression of BdACT with the ellarinacin cluster resulted
413 in formation of a new compound identified as acetyl-ellarinacin, based on its molecular mass signal
414 $[M+H=497.3]$ and the fact that its formation required expression of the entire wheat BGC (Fig. 6d,
415 Supplementary Figs. 33,34). The formation of novel compounds through combining genes from the
416 wheat and *B. distachyon* BGCs demonstrates the promiscuous nature of enzymes encoded by genes
417 within these two clusters.

418

419 **DISCUSSION**

420 Despite the importance of wheat as a food and feed crop, our understanding of the molecules that it
421 produces in response to biotic stress remains limited. Conversely, various phytoalexins and their
422 biosynthetic pathways have been well-characterized in other major cereal crops such as rice, maize, oat
423 and sorghum^{48,49}, and serve as potential targets for crop improvement^{48,50}. Research into specialized

424 metabolism in wheat has until recently been hindered by the lack of a fully assembled genome. The
425 availability of a newly assembled genome coupled with the vast amount of available transcriptomic
426 data now opens up opportunities to deploy genomics-driven approaches for discovery of novel
427 metabolic pathways in wheat, including those implicated in plant defense. Here, utilization of wheat
428 genomic and transcriptomic resources has enabled us to identify previously unknown pathogen-induced
429 biosynthetic pathways for flavonoids, diterpenes and triterpenes. These pathways are driven by sets of
430 genes that are co-localized in the wheat genome, forming six BGCs, including two pairs of
431 homoeologous clusters.

432 We identified a cluster on wheat Chr.5D, BGC 4(5D), encoding a biosynthetic pathway for O-
433 methylated flavonoids. Although flavonoids form a ubiquitous and highly diverse class of compounds
434 in plants, BGC 4(5D) is the only identified and functionally validated flavonoid BGC to date. Further
435 research will be needed for full structural assignment of the pathway product. Additionally, two
436 different types of diterpene-producing clusters (BGC 1(2A/2D) and BGC 2(2B)), were identified on
437 group-2 chromosomes, one of which is syntenic to the rice momilactone cluster. However, the pathways
438 encoded by these syntenic BGCs diverge in their early steps due to differential CPS activities (i.e.,
439 producing *syn*-CPP or normal CPP). Notably, momilactone BGCs were also found in genomes of
440 barnyard grass, *Echinochloa crus-galli*⁵¹ and the bryophyte *Calohypnum plumiforme*⁵², the latter having
441 evolved independently from the momilactone cluster in the grasses⁵³. We did not identify a homologous
442 cluster in the other grass genomes that we analyzed, which included *B. distachyon*, oat, barley and
443 maize. A putative terpene cluster homologous to cluster 2(2B) was however found in *B. distachyon*
444 (Fig. 7a, Supplementary Table 7).

445 Finally, a pathogen-induced cluster (BGC 3(5A/5D)) for a novel isoarborinol-derived
446 triterpenoid, ellarinacin, was found in Chr.5 of the A and D genomes, which is conserved in its wild
447 ancestral species, wild emmer wheat and *A. tauschii*, and is comprised of genes co-opted from sterol
448 primary metabolism. Interestingly, we found the ellarinacin BGC to be more highly induced by non-
449 adapted strains of several fungal pathogens. This cluster may thus form part of a wider set of defense
450 responses found to be actively suppressed in wheat by adapted fungal pathogens, presumably via
451 suppression of plant immune response regulators by pathogen-secreted effector proteins³⁹.

452 Microsynteny and homology searches in other grasses revealed the occurrence of a pathogen-
453 induced cluster in *B. distachyon*, homologous to the ellarinacin cluster in wheat. The pathways encoded
454 by these two clusters diverge from the shared intermediate 19-hydroxy-isoarborinol by
455 neofunctionalization of duplicated CYP51 genes, together with recruitment of additional genes for other
456 enzyme families. Recombinant expression experiments showed that at least some components of these
457 clusters are interchangeable, enabling production of molecules that are not produced by either cluster
458 alone (Fig. 7b), and pointing to the importance of enzyme promiscuity in facilitating chemical
459 diversification.

460 In summary, a genomics-driven approach has enabled us to rapidly identify and characterize
461 novel compounds and biosynthetic pathways in bread wheat. These clusters are highly induced in
462 response to infection by various fungal pathogens and PAMPs, suggesting a broad-spectrum role for
463 these clusters in chemical defense against biotic stresses. Correspondingly, co-expressed genes within
464 these clusters were found to be part of a shared regulatory network that includes various transcription
465 factors predicted to be associated with biotic stress responses. Future work is needed to further
466 understand the interactions and potential contribution of each of these pathways to protection from
467 pathogens in wheat and other grasses, as well to elucidate the regulatory network which governs the
468 expression of these pathways.

469

470 **MATERIALS AND METHODS**

471 **Regulatory network analysis**

472 Target gene-transcription factor interactions and GO term enrichment tables were extracted from a
473 GENIE3-generated wheat regulatory network⁸, available at <https://doi.org/10.5447/ipk/2018/7>.
474 Network visualization was done with Cytoscape v3.8⁵⁴. Genbank accessions for benzoxazinoid pathway
475 genes were retrieved from⁵⁵ and matched with IWGSC gene IDs by BlastN on EnsemblPlants
476 (<http://plants.ensembl.org>). WGCNA and GENIE3-generated regulatory network⁸ were generated using
477 IWGSC RefSeq v1.0 gene models. Other analyses described in this manuscript are based on RefSeq
478 v1.1 gene models.

479 **Co-expression analysis**

480 Co-expression within each cluster was assessed by calculation of the Pearson correlation coefficient (r-
481 val) between the expression of a representative scaffold-forming gene from each cluster (*i.e.*, TPS in
482 clusters 1(2A), 1(2D) and 2(2B), OSC in clusters 3(5A) and 3(5D), and chalcone synthase in cluster
483 4(5D)), and other genes in the cluster, within an RNA-seq dataset including 68 experiments from the
484 wheat-expression.com website^{8,19}.

485 **Pairwise alignment with orthologous clusters in wheat ancestral species**

486 Peptide sequences were extracted from EnsemblPlants (<http://plants.ensembl.org>): *Aegilops tauschii*
487 (Aet_v4.0)⁵⁶, *Triticum turgidum subsp. diccoides* (WEWSeq_v1.0)⁵⁷, *Triticum aestivum* (IWGSC)⁶.
488 Gene models were manually corrected to obtain full coding sequences, and putative protein sequences
489 were aligned using LALIGN (<http://www.ebi.ac.uk>).

490 **Microsynteny analyses**

491 To perform microsynteny analysis and generate figures, a python implementation of
492 MCScan⁵⁸, [https://github.com/tanghaibao/jcvi/wiki/MCscan-\(Python-version\)](https://github.com/tanghaibao/jcvi/wiki/MCscan-(Python-version)), was used. FASTA and
493 GFF3 files were retrieved from EnsemblPlants (<http://plants.ensembl.org>) for chromosomes 5A, 5B and
494 5D of *Triticum aestivum* (IWGSC), 5A and 5D of *Triticum turgidum subsp. diccoides*
495 (WEWSeq_v.1.0) and 5D of *Aegilops tauschii* (Aet_v4.0). MCScan ortholog finding and synteny
496 assignment was run with a c-score of 0.99 and a single iteration. For wheat-rice analysis, wheat
497 *Triticum aestivum*_4.0⁵⁹ and rice IRGSP-1.0⁶⁰ assemblies were used.

498 **Semiquantitative RT-PCR in wheat tissues**

499 Bread wheat cv. ‘Chinese Spring’ plants grown in hydroponic cultures were used for collection of
500 coleoptile, root and root tip (5 mm terminal sections) tissues. Sterilized seeds were transferred to sterile
501 polyacrylamide beads (Scotts) equilibrated with Hoagland’s medium #2. Tissues were harvested after
502 5 days of incubation under controlled conditions (16 h/ 8 h light/dark photoperiod, 23°C). Tissues of
503 stem, inflorescence and leaf infected with mildew (*Blumeria graminis f. sp. tritici* isolate FAL92315)

504 were harvested from 9-week-old plants grown in a greenhouse. Leaf and wounded leaf tissues were
505 harvested from 12-day-old plants grown in a controlled environment (16 h/ 8 h light/dark photoperiod,
506 18°C during daytime, 13°C at night). Wounded leaf tissue was collected 3 h after wounding with forceps.
507 For root tips infected with *Gaeumannomyces graminis* (Take-all), tissues were collected from sterile
508 plants germinated on a fungus-containing substrate. 5 mm terminal root sections were collected 6 d
509 after sowing. For preparation of substrate, fungus was grown in 200 ml PD medium at 22°C/130 rpm
510 for 4 d, mycelium washed 5 times with Hoaglands medium #2 and mixed with 20 ml polyacrylate beads
511 (prepared by addition of 5 gr of cross-linked polyacrylate (Miracle-Gro) to 250 ml of Hoagland's
512 medium #2, equilibration of hydrated beads with 3 x 250 ml of medium and subsequent autoclaving).
513 RNA from all samples was extracted using TRI reagent (Sigma-Aldrich), according to manufacturer's
514 protocol. 5 µg total RNA of each sample was used in 20 µl reverse transcription reactions with
515 Superscript III (Thermo Fisher Scientific), according to manufacturer's protocol. 30-cycle PCR
516 reactions containing 0.2 µl cDNA template in 10 µl total reaction volume were performed and analyzed
517 on 1% agarose gels. Oligonucleotides used are specified in Supplementary Table 8.

518 **Semiquantitative RT-PCR in *Brachypodium distachyon* tissues**

519 Tissues for RT-PCR were sampled from greenhouse-grown *B. distachyon* (Bd21) plants or 2-day-old
520 seedlings grown on a Petri dish in a growth cabinet (28°C, 16 h photoperiod). RNA was extracted using
521 RNeasy plant mini kit (Qiagen), treated with DNase (RQ1, Promega) and used for cDNA library
522 preparation with GoScript reverse transcriptase (Promega), using oligo(dT) primers. All PCR reactions
523 were carried out on Eppendorf Mastercycler pro thermal cycler, for 40 cycles with 55°C annealing
524 temp., using GoTaq G2 Green Master Mix (Promega) and oligonucleotides detailed in Supplementary
525 Table 8. Electrophoresis of PCR products was done on EtBr-stained 1% agarose gels and photographed
526 on a Gel Doc XR instrument (Bio-Rad).

527 **Inoculation of detached wheat leaves with powdery mildew**

528 For gene expression profiling by qRT-PCR, detached leaves from 10-day-old Chinese Spring wheat
529 plants, grown in a growth cabinet (18°C, 16 h day-length under fluorescent lights supplemented with

530 near-UV lights and 12°C for 8 h in the dark), were inoculated with *Blumeria graminis* f. sp. *tritici*
531 (isolate FAL92315, maintained on the susceptible wheat cv. Cerco), or with *Blumeria graminis* f. sp.
532 *hordei* (CH4.8 isolate, maintained on the susceptible barley cv. Golden Promise). Non-inoculated
533 detached leaves kept in same conditions were used as controls. Leaf segments of ~4 cm length were
534 placed in boxes containing water with 0.5% agar and 100 mg L⁻¹ benzimidazole, and were inoculated
535 by blowing fresh spores into settling towers placed over the plant material, according to the method
536 of⁶¹. Following inoculation, plant material was kept in growth cabinet at constant temperature of 15°C
537 and 16 h day-length, and samples collected 12 h and 24 h post-inoculation.

538 **Treatment of detached wheat leaves with elicitors**

539 2-3 cm leaf sections were cut from 1st leaf of 10-day old Chinese Spring seedlings grown in soil. Leaf
540 sections were kept in H₂O in a Petri dish for 24 hours in a 22°C lighted growth cabinet (16 h/ 8 h
541 light/dark photoperiod), then transferred to Petri dishes containing different solutions and kept in same
542 cabinet: 150 µM methyl jasmonate (Sigma-Aldrich), 500 µM salicylic acid, pH 6.0 (Sigma-Aldrich),
543 0.5 mg/ml chitin (NaCoSy) or H₂O. All solutions also contained 0.02% Tween-20. Samples for qRT-
544 PCR analysis were collected after 2 h or 12 h. Four biological replicates of MeJa-treated leaves were
545 collected after three days of treatment for GC-MS analyses.

546 **Treatment of *B. distachyon* with methyl jasmonate**

547 Sections were cut from aerial parts of *B. distachyon* Bd21 plants grown in soil for 2.5 weeks. Samples
548 were kept in H₂O in a Petri dish for 24 h in a 22°C lighted growth cabinet (16 h photoperiod), then
549 transferred to Petri dishes containing 150 µM methyl jasmonate and 0.02% Tween-20, or 0.02% Tween-
550 20 in H₂O, and kept in same cabinet. Four biological replicates of samples were collected for GC-MS
551 analysis after three days.

552 **Quantitative real-time PCR (qRT-PCR) of wheat**

553 For qRT-PCR analysis of wheat leaves inoculated with powdery mildew or treated with elicitors, three
554 biological replicates, each containing three leaf samples, were tested for each time point. RNA was
555 extracted using TRI reagent (Sigma-Aldrich), according to manufacturer's protocol. Following DNase

556 treatment (RQ1, Promega), RNA was reverse-transcribed with M-MLV reverse-transcriptase
557 (ThermoFisher Scientific) using a 1:1 mix of random hexamers and oligo(dT) primers. All
558 oligonucleotides (Supplementary Table 8) were designed using Primer3 software⁶², with at least one
559 homoeolog-specific oligo per each pair used. qRT-PCR was performed on a CFX96 Touch Real-Time
560 PCR instrument (Bio-Rad) in the following conditions: initial step in the thermal cycler for 3 min at
561 95°C, followed by PCR amplification for 40 cycles of 10 s at 95°C and 30 s at 59°C, and finally
562 dissociation analysis to confirm the specificity of PCR products with 0.5°C ramping from 55°C to 95°C.
563 Each 10 µl reaction was comprised of 5 µl LightCycler 480 SYBR Green I Master mix (Roche Life
564 Science), 2 µl cDNA template, 2 µl H₂O and 1 µl primer mix (0.5 µM each primer). Relative transcript
565 levels were calculated according to the Pfaffl method⁶³, using the housekeeping gene β -tubulin (TUBB)
566 as reference⁶⁴.

567 **Quantitative real-time PCR (qRT-PCR) of *Brachypodium distachyon***

568 *B. distachyon* accession Bd3-1 seeds were soaked, peeled, and placed between three filter paper layers
569 soaked in 5 ml water. The seeds were stratified for 5 days at 5°C in the dark and one day at 22°C (16h/8h
570 - light/dark photoperiod) in a controlled environment growth cabinet. For Fusarium root rot (FRR)
571 material, ten stratified seeds were placed on 9 cm² filter square paper on 50 ml 0.8% water agar. All
572 plates were placed in a plant propagator with water-soaked paper towels, angled 20° from the upright
573 position, and stored for 3 days at 22°C (16 h/ 8 h - light/dark photoperiod, variable humidity). *Fusarium*
574 *graminearum* isolate PH1 was maintained on potato dextrose agar (PDA) at 22°C 16 h/ 8 h - light/dark
575 photoperiod in a controlled environment growth cabinet. One 9 cm diameter Petri-dish of seven-day
576 old *F. graminearum* mycelia was blended to a slurry with 1 ml water and applied to three points on
577 each root (root tip, mid root, and near seed) using a 10 ml syringe. The inoculum slurry was removed
578 once infection was visible at 1 dpi and the roots were rinsed with sterile distilled water. Immediately
579 after and then every two days, ten roots per plate (one biological replicate pool) were cut and flash
580 frozen in liquid nitrogen. For Fusarium Head Blight (FHB) material, seeds were sown in 50% peat/sand
581 and 50% John Innes mix 2 (two seeds per 8 cm² pot). Plants were then maintained for six weeks at 22°C
582 (20 h/ 4 h - light/dark photoperiod, 70% humidity) in controlled environment growth cabinet until mid-

583 anthesis. Before the dark period, pots and matting was watered until run-off, spikes were inoculated
584 with 1×10^6 spores/cm² amended with 0.05% Tween-20, and all plants were enclosed in clear plastic
585 bags to maintain high humidity for three days. Immediately after and then every two days, three spikes
586 from different plants were pooled and flash frozen in liquid nitrogen. For conidial suspension inoculum,
587 Mung Bean (MB) broth⁶⁵ with a 1 cm² *F. graminearum* PDA mycelial plug was incubated at 23-25°C,
588 200 rpm for seven days. The inoculum was filtered with cheesecloth and quantified using a
589 haemocytometer.

590 RNA from FHB, FRR, and control samples was extracted using a QIAGEN RNAeasy plant
591 mini kit as per standard protocol. RNA was then immediately cleaned using Turbo DNA-free kits
592 (Invitrogen) as per standard protocol except for two rounds of Turbo DNase treatment. Subsequently
593 cDNA was prepared using SuperScript III Reverse Transcriptase (Thermo Fisher Scientific), as per
594 standard protocol. All oligonucleotides (Supplementary Table 8) were designed using Primer3
595 software⁶². Reverse transcriptase qPCR was performed in a Framestar-480/384 well plate containing 5
596 µl of 2x SYBR Green JumpStart Taq ReadyMix (Sigma-Aldrich), 2 µl cDNA, 0.6 µl of 10 µM per
597 primer, and 1.8 µl water per well. The thermocycling protocol 300 s 95°C, 45x(94°C 10 s, 58°C 10 s,
598 72°C 10 s, 75°C 2 s (single acquisition)), followed by dissociation analysis by ramping from 65°C to
599 97°C, was performed on a Roche LightCycler LC480. Cq values and primer efficiency were quantified
600 using the LinRegPCR tool (Amsterdam UMC Heart Failure Research Centre). Relative quantification
601 was calculated according to the Pfaffl method⁶³, using the housekeeping gene GAPDH as reference.

602 **Generation of DNA constructs**

603 For cloning of *Brachypodium distachyon* and wheat genes, RNA was extracted from mature plant leaves
604 of *B. distachyon* (accession Bd21) or leaves from 10-day-old *Triticum aestivum* plants (Chinese Spring),
605 infected with powdery mildew (*Blumeria graminis f. sp. tritici*), using RNeasy plant mini kit (Qiagen).
606 RNA was treated with RQ1 DNase (Promega) and cDNA libraries prepared with Superscript IV or
607 Superscript III reverse transcriptase kits (Thermo Fisher Scientific), using oligo(dT) primers, according
608 to manufacturer's protocols. TaOMT3, TaOMT6, TaOMT8, TaCYP71C164_5D, TaCYP71F53_5D,
609 BdOSC2, CYP51H13P/H13_5A/14/H15/H16/H35/H37 were amplified from cDNA using Phusion

610 DNA polymerase (Thermo Fisher Scientific) or Q5 DNA polymerase (New England Biolabs).
611 TaOSC_5D, TaHSD, AsOSC1, AsCYP51H73 were synthesized by General Biosystems, Durham, NC,
612 USA. BdACT, BdMeTr, TaCHS1, chi-1D, TaCPS-D2, TaKSL-D1 and *Taxus canadensis* GGPPS were
613 synthesized by Twist Bioscience, San Francisco, CA, USA. TaCPS-D2, TaKSL-D1 and TcGGPPS lack
614 signal sequences, to allow for cytosolic localization in *N. benthamiana* expression⁶⁶. *A. tauschii* IAS
615 coding sequence was derived from TaOSC_5D sequence with site directed mutagenesis⁶⁷ to obtain a
616 single mutation (I581S). Synthesized and cDNA-amplified genes from triterpene and diterpene BGCs
617 were cloned into a pCAMBIA-based⁶⁸ plant expression vector with Goldenbraid cloning⁶⁹, using BsaI
618 and BsmBI (New England Biolabs) and T4 DNA ligase enzymes (New England Biolabs). Gene
619 expression in final vectors is driven by *Solanum lycopersicum* ubiquitin 10 promoter and terminator⁷⁰.
620 Synthesized and cDNA-amplified genes from flavonoid cluster were cloned into a pDONR207 Gateway
621 entry vector and subcloned into a pEAQ-HT-DEST1 plasmid⁷¹ using BP and LR clonase enzyme mixes
622 (Thermo Fisher Scientific), respectively. Full coding sequences of all synthesized and PCR-cloned
623 genes used in this study are found in Supplementary Methods. Oligonucleotides used for amplification
624 and sub-cloning are specified in Supplementary Table 8.

625 **Agroinfiltration-mediated transient expression in *N. benthamiana***

626 Plant expression vectors were transformed into *Agrobacterium tumefaciens* GV3101 via
627 electroporation. Agrobacteria cultures were grown overnight in 28°C in LB media and resuspended in
628 MMA buffer (10 mM MgCl₂, 10 mM MES pH 5.6, 100 µM acetosyringone) to O.D.₆₀₀ 0.2. For co-
629 expression of several genes, O.D.₆₀₀ 0.2 cultures of strains expressing different genes were mixed 1:1
630 prior to infiltration. Cultures were infiltrated by syringe into leaves of 5 weeks old greenhouse-grown
631 *N. benthamiana* plants. The plants were further maintained in the greenhouse after infiltration.
632 Infiltrated leaves were harvested 5 days post infection, freeze-dried and ground.

633 **Metabolite extraction from agroinfiltrated *N. benthamiana* leaves for GC-MS analyses**

634 Diterpenes: for analysis of TaCPS-D2 and TaKSL-D1 transient expression, 5 mg of *N. benthamiana*
635 leaf samples were extracted in 850 µl ethyl acetate for 1 h in room temperature, with agitation.

636 Following removal of plant tissue by centrifugation, 750 μ l from each extract was evaporated and
637 reconstituted in 75 μ l ethyl acetate. Triterpenes: for analysis of wheat BGC 3(5D) genes expression, 5
638 mg samples were extracted in 500 μ l ethyl acetate with 5 μ g/ml 5 α -cholestan-3 β -ol. For analysis of *B.*
639 *distachyon* brachynacin cluster genes expression, 5 mg samples were extracted in 500 μ l methanol with
640 5 μ g/ml 5 α -cholestan-3 β -ol. For analysis of oat and *A. tauschii* OSC and CYP51 genes, 5 mg samples
641 were extracted in 500 μ l ethyl acetate. For analysis of combined expression of wheat BGC 3(5D) genes
642 and BdACT, 5 mg samples were extracted in 300 μ l ethyl acetate. For analysis of combined expression
643 of *B. distachyon* brachynacin cluster genes and TaHSD, 5 mg samples were extracted in 500 μ l ethyl
644 acetate with 5 μ g/ml 5 α -cholestan-3 β -ol. All triterpene extractions from *N. benthamiana* leaves were
645 done in room temperature for 1 hour, with agitation. For all triterpene extractions, following the removal
646 of plant tissue by centrifugation, 200 μ l were evaporated and reconstituted in 70 μ l TMS+pyridine
647 (Sigma-Aldrich). Samples were derivatized for 0.5 h in 70°C.

648 **Metabolite extraction from MeJA-treated wheat and *B. distachyon* leaves for GC-MS analyses**

649 MeJA-treated wheat leaf sections were freeze-dried and ground. 25 mg from each sample were extracted
650 in 800 μ l ethyl acetate containing 5 μ g/ml 5 α -cholestan-3 β -ol, with agitation for 2 h in 40°C. Following
651 removal of tissue by centrifugation and filtration with 0.22 μ l filter mini columns (Norgen), 700 μ l from
652 each extract was evaporated and reconstituted in 70 μ l TMS with pyridine (Sigma-Aldrich). Samples
653 were derivatized for 0.5 h in 70°C. MeJA-treated *B. distachyon* leaf sections were freeze-dried and
654 ground. 25 mg from each ground sample were extracted in 1100 μ l methanol containing 2.5 μ g/ml 5 α -
655 cholestan-3 β -ol, with agitation for 2 h in 40°C. Following removal of tissue by centrifugation and
656 filtration, 800 μ l from each extract was evaporated and reconstituted in 70 μ l TMS with pyridine
657 (Sigma-Aldrich). Samples were derivatized for 0.5 h in 70°C.

658 **GC-MS analysis of diterpenes and triterpenes from *N. benthamiana* and grasses leaf extracts**

659 GC-MS analysis was performed using an Agilent 7890B instrument with a Zebron ZB5-HT Inferno
660 column (Phenomenex). For triterpenes analysis, a previously described method³⁰ was used: injections
661 were performed in pulsed splitless mode (30 psi pulse pressure). Inlet temperature was set to 250°C.

662 GC oven temperature was initially held at 170°C for 2 mins, subsequently ramped to 300°C at 20°C/min
663 and held at 300°C for an additional 11.5 min (20 min total run time). The GC oven was coupled to an
664 Agilent 5977B MS detector set to scan mode, from 60 to 800 mass units (solvent delay 8 min). For
665 semi-quantification of brachynacin in *B. distachyon* leaves, Selected Ion Monitoring (SIM) mode was
666 used, for detection of brachynacin (m/z 170.1, 340.2, 387.3, 400.3, 445.4, 475.4, 500.4 ions were
667 monitored) and internal standard 5 α -cholestan-3 β -ol (m/z 215.1, 355.4, 445.5, 460.5 ions were
668 monitored), with 100 ms dwell time for each ion. Diterpenes analysis was based on a previously
669 described method⁶⁶: injections were performed in splitless mode. Inlet temperature was set to 280°C.
670 GC oven temperature was initially held at 130°C for 2 mins, subsequently ramped up to 250°C at
671 8°C/min, followed by ramping up to 310°C at 10°C/min and held at 310°C for an additional 5 min
672 (28 min total run time). The MS detector was set to scan mode, from 50 to 550 mass units (solvent delay
673 4 min).

674 **Metabolite extraction from agroinfiltrated *N. benthamiana* leaves for LC-MS analyses**

675 For LC-MS analysis of recombinantly expressed wheat BGC 3(5D) genes (including combined
676 expression with BdACT), 25 mg of each sample were extracted in 2 ml methanol in room temperature
677 for 1 h, with agitation. Following removal of plant tissue by centrifugation, extracts were partitioned
678 twice with 2 ml hexane and filtered with 0.22 μ l filter mini columns (Norgen). Extracts were evaporated
679 and resuspended in 100 μ l methanol. For analysis of recombinantly expressed *B. distachyon*
680 brachynacin cluster genes (including combined expression with TaHSD), 10 mg of each sample were
681 extracted in 400 μ l 80% methanol in room temperature for 1 h, with agitation. Following removal of
682 plant tissue by centrifugation, extracts were partitioned with 500 μ l hexane and filtered. Extracts were
683 evaporated and resuspended in 100 μ l 80% methanol. For analysis of wheat flavonoid cluster genes,
684 250 mg freeze-dried and ground samples were extracted with 4 mL methanol at room temperature for
685 1 h. Extracts were fully evaporated, resuspended in 200 μ L methanol, and filtrated through a mini
686 column (pore size 0.22 μ m, Geneflow). Filtered samples were transferred to glass autosampler vials
687 and 20 μ L of each sample was analyzed by UHPLC-CAD-PDA-MS.

688 **LC-MS analyses of triterpenes from *N. benthamiana* leaf extracts**

689 Leaf extracts were analyzed by reverse phase HPLC on a Shimadzu LCMS-2020 single quadrupole
690 mass spectrometer coupled with a Dionex Corona Veo RS charged aerosol detector (Thermo Scientific),
691 using a Kinetex 2.6 μm XB-C18 100 Å, 50 x 2.1 mm LC Column (Phenomenex). MS data was collected
692 using combined electrospray ionization (ESI) and atmospheric pressure chemical ionization (APCI) in
693 positive mode. 10 μl samples were injected using 12 min, 14.5 min or 30 min mobile phase gradient
694 methods using solvent A- water with 0.1% formic acid and solvent B- methanol with 0.1% formic acid,
695 as follows. 12 min method: 50% B hold from 0 to 0.75 min, 50% to 90% B from 0.75 to 8 min, 90% B
696 hold from 8 to 10 min, 90% to 50% B from 10 to 10.5 min, 50% B hold from 10.5 to 12 min. Flow rate,
697 0.4 ml/min. MS scan, m/z 250 – 1900. Column oven temperature, 40°C. 14.5 min method: 70% to 95%
698 B from 0 to 10 min, 95% B hold from 10 to 11 min, 95% to 70% B from 11 to 11.1 min, 70% B hold
699 from 11.1 to 14.5 min. Flow rate, 0.5 ml/min. MS scan, m/z 200 – 1200. Column oven temperature,
700 30°C. 30 min method: 15% B hold from 0 to 0.15 min, 15% to 60% B from 0.15 to 26 min, 60% to
701 100% B from 26 to 26.5 min, 100% B hold from 26.5 to 28.5 min, 100% to 15% B from 28.5 to 29 min,
702 15% B hold from 29 to 30 min. Flow rate, 0.3 ml/min. MS scan, m/z 100 – 1500. Column oven
703 temperature, 30°C.

704 **LC-MS analyses of flavonoids from *N. benthamiana* leaf extracts**

705 Leaf extracts were analyzed by reverse phase HPLC on a Shimadzu LCMS-2020 single quadrupole
706 mass spectrometer coupled with a Dionex Corona Veo RS charged aerosol detector (Thermo Scientific)
707 and a SPD-M20A HPLC Photodiode Array Detector (PDA; Shimadzu), using a Kinetex 2.6 μm XB-
708 C18 100 Å, 50 x 2.1 mm LC Column (Phenomenex), kept at 30°C. Water containing 0.1% formic acid
709 (FA) and acetonitrile containing 0.1% formic acid (FA) were used as mobile phases A and B,
710 respectively, with a flow rate of 0.2 mL/min. A gradient elution program was applied as follows: 0-1.5
711 min linearly increased from 0% to 10% B, 1.5-26 min linearly increased from 10% to 60% B, 26-26.5
712 min linearly increased from 60% to 80% B, 26.5-28.5 min linearly increased from 80% to 100% B,
713 28.5-29 min linearly decreased from 100% to 10% B hold on for another 1 min for re-equilibration,
714 giving a total run time 30 min. MS detection was performed in both positive and negative ESI range of
715 m/z 50–1500 with the following settings: desolvation temperature was 250°C; drying gas flow, 15

716 L/min; detector voltage was 1.25 kV; and nebulizing gas flow, 1.5 L/min. PDA chromatograms were
717 recorded in a 200–600 nm range using a deuterium (D2) and tungsten (W) light source.

718 High-resolution mass spectrometry analysis of the metabolites was carried out on a Q Exactive
719 instrument (Thermo Scientific). Chromatography was performed using a Kinetex 2.6 μm XB-C18 100
720 \AA , 50 mm x 2.1 mm (Phenomenex) column kept at 30°C. Water containing 0.1% formic acid (FA) and
721 acetonitrile containing 0.1% formic acid (FA) were used as mobile phases A and B, respectively with a
722 flow rate of 0.4 mL/min. A gradient elution program was applied as follows: 0-0.75 min linearly
723 increased from 0% to 10% B, 0.75-13 min linearly increased from 10% to 60% B, 13-13.25 min linearly
724 increased from 60% to 80% B, 13.25-14.25 min linearly increased from 80% to 100% B, 14.25-14.5
725 min linearly decreased from 100% to 10% B hold on for another 2.5 min for re-equilibration, giving a
726 total run time 17 min. MS detection was performed in both positive and negative ESI range of 100–
727 1500 m/z . The analysis of the 3D field of the Photodiode-Array Detection (PDA) was recorded in a
728 200–600 nm range using a vanquish detector (Thermo Scientific).

729 **Large-scale agroinfiltration, extraction and purification of triterpenoids**

730 Vacuum-mediated large-scale agroinfiltrations of *N. benthamiana* plants, and downstream extraction
731 and purification of triterpenoid products were based on a previously described method^{30,72}. Specific
732 methods for extraction and purification of the metabolites are detailed in Supplementary Methods.

733 **General considerations for NMR.**

734 NMR spectra were recorded in Fourier transform mode at a nominal frequency of 600 MHz for ^1H
735 NMR, and 150 MHz for ^{13}C NMR (unless specified otherwise), using the specified deuterated solvent.
736 Chemical shifts were recorded in ppm and referenced to the residual solvent peak or to an internal TMS
737 standard. Multiplicities are described as, s = singlet, d = doublet, dd = doublet of doublets, dt = doublet
738 of triplets, t = triplet, q = quartet, quint = quintet, tquin = triplet of quintets, m = multiplet, br = broad,
739 appt = apparent; coupling constants are reported in hertz as observed and not corrected for second order
740 effects.

741 **Genomic positioning of wheat BGC homologs in other grasses**

742 Protein sequences of all co-expressed genes from wheat BGCs 1(2D), 2(2B), 3(5D) and 4(5D) were
743 used as BlastP queries against the following genome assemblies: *Zea mays* B73 RefGen_v4⁷³, *Hordeum*
744 *vulgare* cv. Morex r1⁷⁴, *Brachypodium distachyon* Bd21 v3.1⁴⁴, *Oryza sativa ssp. japonica*
745 cv. Nipponbare v7.0⁷⁵ and *Avena strigosa* S75 v2.0⁴³. BlastP searches in all assemblies except *Avena*
746 *strigosa* were performed in Phytozome13 (<https://phytozome-next.jgi.doe.gov/>)⁴⁵, using default
747 parameters. Genomic locations of top BlastP hits in each species were visualized using Circos software
748 v0.69-9⁷⁶.

749

750 ACKNOWLEDGEMENTS

751 We would like to thank Paul Brett, Lionel Hill, Amr El-Demerdash, James Reed, Hannah Hodgson,
752 Rebecca Casson and Sergey Nepogodiev for assistance and advice on analytical chemistry analyses,
753 Andrew Steed and Rachel Burns for assistance with wheat and *B. distachyon* pathogen infections,
754 Nikolai Adamski for helpful discussions, JIC horticultural services staff for assistance with plant
755 cultivation, Noam Chayut and Simon Orford (JIC Germplasm Resource Unit) for providing ‘Chinese
756 Spring’ seeds, and Christine Faulkner for providing chitin. *Blumeria graminis* isolates CH4.8
757 and FAL92315 were kindly provided by Lesley Boyd (NIAB). G.P. is supported by a Royal Society
758 Kohn International Fellowship (NIFAR1\180677) and a Marie Skłodowska-Curie Individual Fellowship
759 (838242). A.O.’s lab is supported by the Biological Sciences Research Council (BBSRC)-funded
760 Institute Strategic Programme Grant ‘Molecules from Nature’ (BB/P012523/1) and the John Innes
761 Foundation.

762

763 AUTHOR CONTRIBUTIONS

764 G.P, M.D and A.O conceived and designed the experiments. G.P, M.D, R.C.M, J.H and L.C
765 performed the experiments. G.P, M.J.S and C.O analyzed the data. R.R.G, H.S, P.B and D.R.N

766 advised, analyzed and contributed data. J.B, P.N, C.U and A.O jointly supervised research. G.P
767 and A.O wrote the manuscript, with contributions from all authors.

768

769 **COMPETING INTERESTS STATEMENT**

770 The authors declare no competing interests.

771

REFERENCES

- 772 1. FAOSTAT (2019) (<http://www.fao.org/faostat/en/#data/QC>).
- 773 2. Savary, S. *et al.* The global burden of pathogens and pests on major food crops. *Nat Ecol Evol*
774 **3**, 430-439 (2019).
- 775 3. Deutsch, C.A. *et al.* Increase in crop losses to insect pests in a warming climate. *Science* **361**,
776 916-919 (2018).
- 777 4. Delgado-Baquerizo, M. *et al.* The proportion of soil-borne pathogens increases with warming
778 at the global scale. *Nature Climate Change* **10**, 550-+ (2020).
- 779 5. Krattinger, S.G. & Keller, B. Molecular genetics and evolution of disease resistance in
780 cereals. *New Phytol* **212**, 320-32 (2016).
- 781 6. Appels, R. *et al.* Shifting the limits in wheat research and breeding using a fully annotated
782 reference genome. *Science* **361**(2018).
- 783 7. Polturak, G. & Osbourn, A. The emerging role of biosynthetic gene clusters in plant defense
784 and plant interactions. *PLoS Pathog* **17**, e1009698 (2021).
- 785 8. Ramirez-Gonzalez, R.H. *et al.* The transcriptional landscape of polyploid wheat. *Science* **361**,
786 662-+ (2018).
- 787 9. Ahmed, S. & Kovicich, N. Regulation of phytoalexin biosynthesis for agriculture and human
788 health. *Phytochemistry Reviews* **20**, 483-505 (2021).

- 789 10. Kumar, M. *et al.* Heat Shock Factors HsfB1 and HsfB2b Are Involved in the Regulation of
790 Pdf1.2 Expression and Pathogen Resistance in Arabidopsis. *Molecular Plant* **2**, 152-165
791 (2009).
- 792 11. Nomura, T., Ishihara, A., Yanagita, R.C., Endo, T.R. & Iwamura, H. Three genomes
793 differentially contribute to the biosynthesis of benzoxazinones in hexaploid wheat.
794 *Proceedings of the National Academy of Sciences of the United States of America* **102**, 16490-
795 16495 (2005).
- 796 12. Zhou, S., Richter, A. & Jander, G. Beyond Defense: Multiple Functions of Benzoxazinoids in
797 Maize Metabolism. *Plant Cell Physiol* **59**, 1528-1537 (2018).
- 798 13. Piasecka, A., Jedrzejczak-Rey, N. & Bednarek, P. Secondary metabolites in plant innate
799 immunity: conserved function of divergent chemicals. *New Phytol* **206**, 948-64 (2015).
- 800 14. Osbourn, A. Secondary metabolic gene clusters: evolutionary toolkits for chemical
801 innovation. *Trends Genet* **26**, 449-57 (2010).
- 802 15. Treutter, D. Significance of flavonoids in plant resistance and enhancement of their
803 biosynthesis. *Plant Biology* **7**, 581-591 (2005).
- 804 16. Thimmappa, R., Geisler, K., Louveau, T., O'Maille, P. & Osbourn, A. Triterpene biosynthesis
805 in plants. *Annu Rev Plant Biol* **65**, 225-57 (2014).
- 806 17. Schmelz, E.A. *et al.* Biosynthesis, elicitation and roles of monocot terpenoid phytoalexins.
807 *Plant Journal* **79**, 659-678 (2014).
- 808 18. Murphy, K.M. & Zerbe, P. Specialized diterpenoid metabolism in monocot crops:
809 Biosynthesis and chemical diversity. *Phytochemistry* **172**(2020).
- 810 19. Borrill, P., Ramirez-Gonzalez, R. & Uauy, C. expVIP: a Customizable RNA-seq Data
811 Analysis and Visualization Platform. *Plant Physiology* **170**, 2172-2186 (2016).
- 812 20. Goodin, M.M., Zaitlin, D., Naidu, R.A. & Lommel, S.A. *Nicotiana benthamiana*: its history
813 and future as a model for plant-pathogen interactions. *Mol Plant Microbe Interact* **21**, 1015-
814 26 (2008).

- 815 21. Wilderman, P.R., Xu, M.M., Jin, Y.H., Coates, R.M. & Peters, R.J. Identification of syn-
816 pimara-7,15-diene synthase reveals functional clustering of terpene synthases involved in rice
817 phytoalexin/allelochemical biosynthesis. *Plant Physiology* **135**, 2098-2105 (2004).
- 818 22. Shimura, K. *et al.* Identification of a biosynthetic gene cluster in rice for momilactones.
819 *Journal of Biological Chemistry* **282**, 34013-34018 (2007).
- 820 23. Swaminathan, S., Morrone, D., Wang, Q., Fulton, D.B. & Peters, R.J. CYP76M7 Is an ent-
821 Cassadiene C11 alpha-Hydroxylase Defining a Second Multifunctional Diterpenoid
822 Biosynthetic Gene Cluster in Rice. *Plant Cell* **21**, 3315-3325 (2009).
- 823 24. Peters, R.J. Two rings in them all: the labdane-related diterpenoids. *Nat Prod Rep* **27**, 1521-
824 30 (2010).
- 825 25. Wu, Y. *et al.* Functional characterization of wheat copalyl diphosphate synthases sheds light
826 on the early evolution of labdane-related diterpenoid metabolism in the cereals.
827 *Phytochemistry* **84**, 40-6 (2012).
- 828 26. Zhou, K. *et al.* Functional characterization of wheat ent-kaurene(-like) synthases indicates
829 continuing evolution of labdane-related diterpenoid metabolism in the cereals.
830 *Phytochemistry* **84**, 47-55 (2012).
- 831 27. Toyomasu, T. *et al.* Cloning and characterization of cDNAs encoding ent-copalyl diphosphate
832 synthases in wheat: insight into the evolution of rice phytoalexin biosynthetic genes. *Biosci*
833 *Biotechnol Biochem* **73**, 772-5 (2009).
- 834 28. Ahn, S., Anderson, J.A., Sorrells, M.E. & Tanksley, S.D. Homoeologous relationships of rice,
835 wheat and maize chromosomes. *Mol Gen Genet* **241**, 483-90 (1993).
- 836 29. Kitaoka, N., Wu, Y., Zi, J. & Peters, R.J. Investigating inducible short-chain alcohol
837 dehydrogenases/reductases clarifies rice oryzalexin biosynthesis. *Plant J* **88**, 271-279 (2016).
- 838 30. Reed, J. *et al.* A translational synthetic biology platform for rapid access to gram-scale
839 quantities of novel drug-like molecules. *Metab Eng* **42**, 185-193 (2017).
- 840 31. Geisler, K. *et al.* Biochemical analysis of a multifunctional cytochrome P450 (CYP51)
841 enzyme required for synthesis of antimicrobial triterpenes in plants. *Proceedings of the*
842 *National Academy of Sciences of the United States of America* **110**, E3360-E3367 (2013).

- 843 32. Rahier, A., Darnet, S., Bouvier, F., Camara, B. & Bard, M. Molecular and enzymatic
844 characterizations of novel bifunctional 3 beta-hydroxysteroid dehydrogenases/C-4
845 decarboxylases from *Arabidopsis thaliana*. *Journal of Biological Chemistry* **281**, 27264-
846 27277 (2006).
- 847 33. Herl, V., Frankenstein, J., Meitingner, N., Muller-Uri, F. & Kreis, W. A(5)-3 beta-
848 hydroxysteroid dehydrogenase (3 beta HSD) from *Digitalis lanata*. Heterologous expression
849 and characterisation of the recombinant enzyme. *Planta Medica* **73**, 704-710 (2007).
- 850 34. Sonawane, P.D. *et al.* Short-chain dehydrogenase/reductase governs steroidal specialized
851 metabolites structural diversity and toxicity in the genus *Solanum*. *Proceedings of the*
852 *National Academy of Sciences of the United States of America* **115**, E5419-E5428 (2018).
- 853 35. Huang, A.C.C. *et al.* A specialized metabolic network selectively modulates *Arabidopsis* root
854 microbiota. *Science* **364**, 546-+ (2019).
- 855 36. Moummou, H., Kallberg, Y., Tonfack, L.B., Persson, B. & van der Rest, B. The Plant Short-
856 Chain Dehydrogenase (SDR) superfamily: genome-wide inventory and diversification
857 patterns. *Bmc Plant Biology* **12**(2012).
- 858 37. Cabello-Hurtado, F. *et al.* Optimized expression and catalytic properties of a wheat
859 obtusifoliol 14 alpha-demethylase (CYP51) expressed in yeast - Complementation of erg11
860 Delta yeast mutants by plant CYP51. *European Journal of Biochemistry* **262**, 435-446 (1999).
- 861 38. Qi, X. *et al.* A different function for a member of an ancient and highly conserved
862 cytochrome P450 family: From essential sterols to plant defense. *Proceedings of the National*
863 *Academy of Sciences of the United States of America* **103**, 18848-18853 (2006).
- 864 39. Dobon, A., Bunting, D.C.E., Cabrera-Quio, L.E., Uauy, C. & Saunders, D.G.O. The host-
865 pathogen interaction between wheat and yellow rust induces temporally coordinated waves of
866 gene expression. *Bmc Genomics* **17**(2016).
- 867 40. Delventhal, R. *et al.* A comparative analysis of nonhost resistance across the two Triticeae
868 crop species wheat and barley. *Bmc Plant Biology* **17**(2017).
- 869 41. Palmer, M.A. & Bowden, B.N. Variations in sterol and triterpene contents of developing
870 sorghum-bicolor grains. *Phytochemistry* **16**, 459-463 (1977).

- 871 42. Xue, Z.Y. *et al.* Divergent evolution of oxidosqualene cyclases in plants. *New Phytologist*
872 **193**, 1022-1038 (2012).
- 873 43. Li, Y. *et al.* Subtelomeric assembly of a multi-gene pathway for antimicrobial defense
874 compounds in cereals. *Nature Communications* **12**(2021).
- 875 44. Vogel, J.P. *et al.* Genome sequencing and analysis of the model grass *Brachypodium*
876 *distachyon*. *Nature* **463**, 763-768 (2010).
- 877 45. Goodstein, D.M. *et al.* Phytozome: a comparative platform for green plant genomics. *Nucleic*
878 *Acids Research* **40**, D1178-D1186 (2012).
- 879 46. Sibout, R. *et al.* Expression atlas and comparative coexpression network analyses reveal
880 important genes involved in the formation of lignified cell wall in *Brachypodium distachyon*.
881 *New Phytologist* **215**, 1009-1025 (2017).
- 882 47. Mutwil, M. *et al.* PlaNet: Combined Sequence and Expression Comparisons across Plant
883 Networks Derived from Seven Species. *Plant Cell* **23**, 895-910 (2011).
- 884 48. Ahuja, I., Kissen, R. & Bones, A.M. Phytoalexins in defense against pathogens. *Trends in*
885 *Plant Science* **17**, 73-90 (2012).
- 886 49. Ejike, C., Gong, M. & Udenigwe, C.C. Phytoalexins from the Poaceae: Biosynthesis, function
887 and prospects in food preservation. *Food Research International* **52**, 167-177 (2013).
- 888 50. Großkinsky, D.K., van der Graaff, E. & Roitsch, T. Phytoalexin transgenics in crop
889 protection--fairy tale with a happy end? *Plant Sci* **195**, 54-70 (2012).
- 890 51. Guo, L. *et al.* *Echinochloa crus-galli* genome analysis provides insight into its adaptation and
891 invasiveness as a weed. *Nat Commun* **8**, 1031 (2017).
- 892 52. Mao, L.F. *et al.* Genomic evidence for convergent evolution of gene clusters for momilactone
893 biosynthesis in land plants. *Proceedings of the National Academy of Sciences of the United*
894 *States of America* **117**, 12472-12480 (2020).
- 895 53. Zhang, J. & Peters, R.J. Why are momilactones always associated with biosynthetic gene
896 clusters in plants? *Proc Natl Acad Sci U S A* **117**, 13867-13869 (2020).
- 897 54. Shannon, P. *et al.* Cytoscape: A software environment for integrated models of biomolecular
898 interaction networks. *Genome Research* **13**, 2498-2504 (2003).

- 899 55. Makowska, B., Bakera, B. & Rakoczy-Trojanowska, M. The genetic background of
900 benzoxazinoid biosynthesis in cereals. *Acta Physiologiae Plantarum* **37**(2015).
- 901 56. Luo, M.C. *et al.* Genome sequence of the progenitor of the wheat D genome *Aegilops*
902 *tauschii*. *Nature* **551**, 498-+ (2017).
- 903 57. Avni, R. *et al.* Wild emmer genome architecture and diversity elucidate wheat evolution and
904 domestication. *Science* **357**, 93-97 (2017).
- 905 58. Tang, H. *et al.* Unraveling ancient hexaploidy through multiply-aligned angiosperm gene
906 maps. *Genome Res* **18**, 1944-54 (2008).
- 907 59. Alonge, M., Shumate, A., Puiu, D., Zimin, A.V. & Salzberg, S.L. Chromosome-Scale
908 Assembly of the Bread Wheat Genome Reveals Thousands of Additional Gene Copies.
909 *Genetics* **216**, 599-608 (2020).
- 910 60. Kawahara, Y. *et al.* Improvement of the *Oryza sativa* Nipponbare reference genome using
911 next generation sequence and optical map data. *Rice (N Y)* **6**, 4 (2013).
- 912 61. Brown, J.K.M. & Wolfe, M.S. Structure and evolution of a population of *Erysiphe-graminis* f
913 *sp hordei*. *Plant Pathology* **39**, 376-390 (1990).
- 914 62. Untergasser, A. *et al.* Primer3-new capabilities and interfaces. *Nucleic Acids Research*
915 **40**(2012).
- 916 63. Pfaffl, M.W. A new mathematical model for relative quantification in real-time RT-PCR.
917 *Nucleic Acids Research* **29**(2001).
- 918 64. Scholtz, J.J. & Visser, B. Reference gene selection for qPCR gene expression analysis of rust-
919 infected wheat. *Physiological and Molecular Plant Pathology* **81**, 22-25 (2013).
- 920 65. Makandar, R., Essig, J.S., Schapaugh, M.A., Trick, H.N. & Shah, J. Genetically engineered
921 resistance to *Fusarium* head blight in wheat by expression of *Arabidopsis* NPR1. *Mol Plant*
922 *Microbe Interact* **19**, 123-9 (2006).
- 923 66. De La Peña, R. & Sattely, E.S. Rerouting plant terpene biosynthesis enables momilactone
924 pathway elucidation. *Nat Chem Biol* **17**, 205-212 (2021).
- 925 67. Zheng, L., Baumann, U. & Reymond, J.L. An efficient one-step site-directed and site-
926 saturation mutagenesis protocol. *Nucleic Acids Res* **32**, e115 (2004).

- 927 68. Roberts, C. *et al.* A Comprehensive Set of Modular Vectors for Advanced Manipulations and
928 Efficient Transformation of Plants. in *pCAMBIA Vector Release Manual Rockefeller*
929 *Foundation Meeting of the International Program on Rice Biotechnology, September 15-19,*
930 *Malacca, Malaysia (1997).*
- 931 69. Sarrion-Perdigones, A. *et al.* GoldenBraid 2.0: A Comprehensive DNA Assembly Framework
932 for Plant Synthetic Biology. *Plant Physiology* **162**, 1618-1631 (2013).
- 933 70. Dahan-Meir, T. *et al.* Efficient in planta gene targeting in tomato using geminiviral replicons
934 and the CRISPR/Cas9 system. *Plant Journal* **95**, 5-16 (2018).
- 935 71. Sainsbury, F., Thuenemann, E.C. & Lomonosoff, G.P. pEAQ: versatile expression vectors
936 for easy and quick transient expression of heterologous proteins in plants. *Plant*
937 *Biotechnology Journal* **7**, 682-693 (2009).
- 938 72. Stephenson, M.J., Reed, J., Brouwer, B. & Osbourn, A. Transient Expression in Nicotiana
939 Benthamiana Leaves for Triterpene Production at a Preparative Scale. *Jove-Journal of*
940 *Visualized Experiments* (2018).
- 941 73. Schnable, P.S. *et al.* The B73 Maize Genome: Complexity, Diversity, and Dynamics. *Science*
942 **326**, 1112-1115 (2009).
- 943 74. Beier, S. *et al.* Construction of a map-based reference genome sequence for barley, *Hordeum*
944 *vulgare* L. *Sci Data* **4**, 170044 (2017).
- 945 75. Ouyang, S. *et al.* The TIGR Rice Genome Annotation Resource: Improvements and new
946 features. *Nucleic Acids Research* **35**, D883-D887 (2007).
- 947 76. Krzywinski, M. *et al.* Circos: An information aesthetic for comparative genomics. *Genome*
948 *Research* **19**, 1639-1645 (2009).
- 949 77. Ma, J. *et al.* Transcriptome and allele specificity associated with a 3BL locus for Fusarium
950 crown rot resistance in bread wheat. *PLoS One* **9**, e113309 (2014).
- 951 78. Powell, J.J. *et al.* The Fusarium crown rot pathogen *Fusarium pseudograminearum* triggers a
952 suite of transcriptional and metabolic changes in bread wheat (*Triticum aestivum* L.). *Ann Bot*
953 **119**, 853-867 (2017).

- 954 79. Rudd, J.J. *et al.* Transcriptome and Metabolite Profiling of the Infection Cycle of
955 *Zymoseptoria tritici* on Wheat Reveals a Biphasic Interaction with Plant Immunity Involving
956 Differential Pathogen Chromosomal Contributions and a Variation on the Hemibiotrophic
957 Lifestyle Definition. *Plant Physiology* **167**, 1158-+ (2015).
- 958 80. Zhang, H. *et al.* Large-scale transcriptome comparison reveals distinct gene activations in
959 wheat responding to stripe rust and powdery mildew. *BMC Genomics* **15**, 898 (2014).

960

961

962

963

964

965

966

967

968

969

970

971

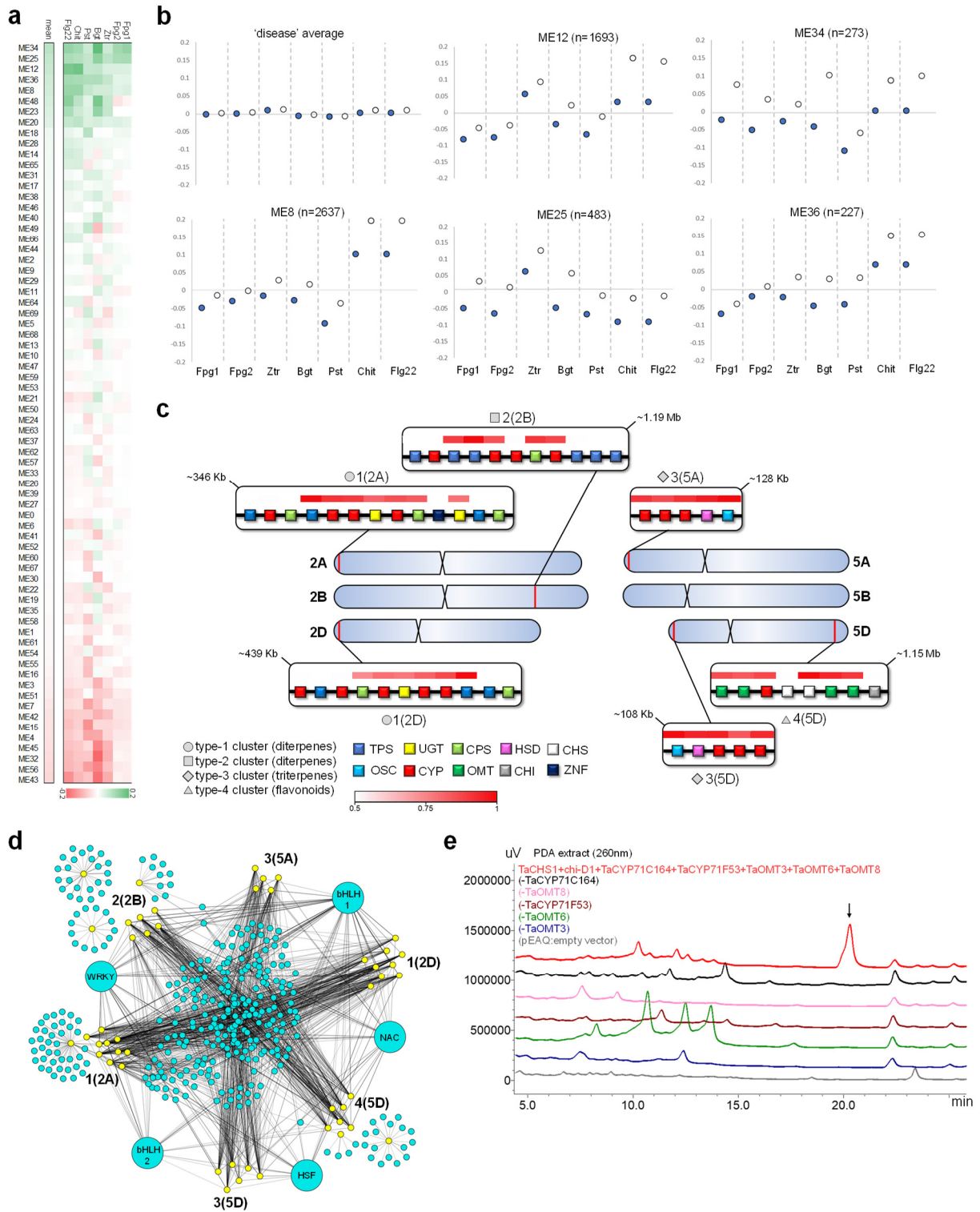
972

973

974

975 FIGURES

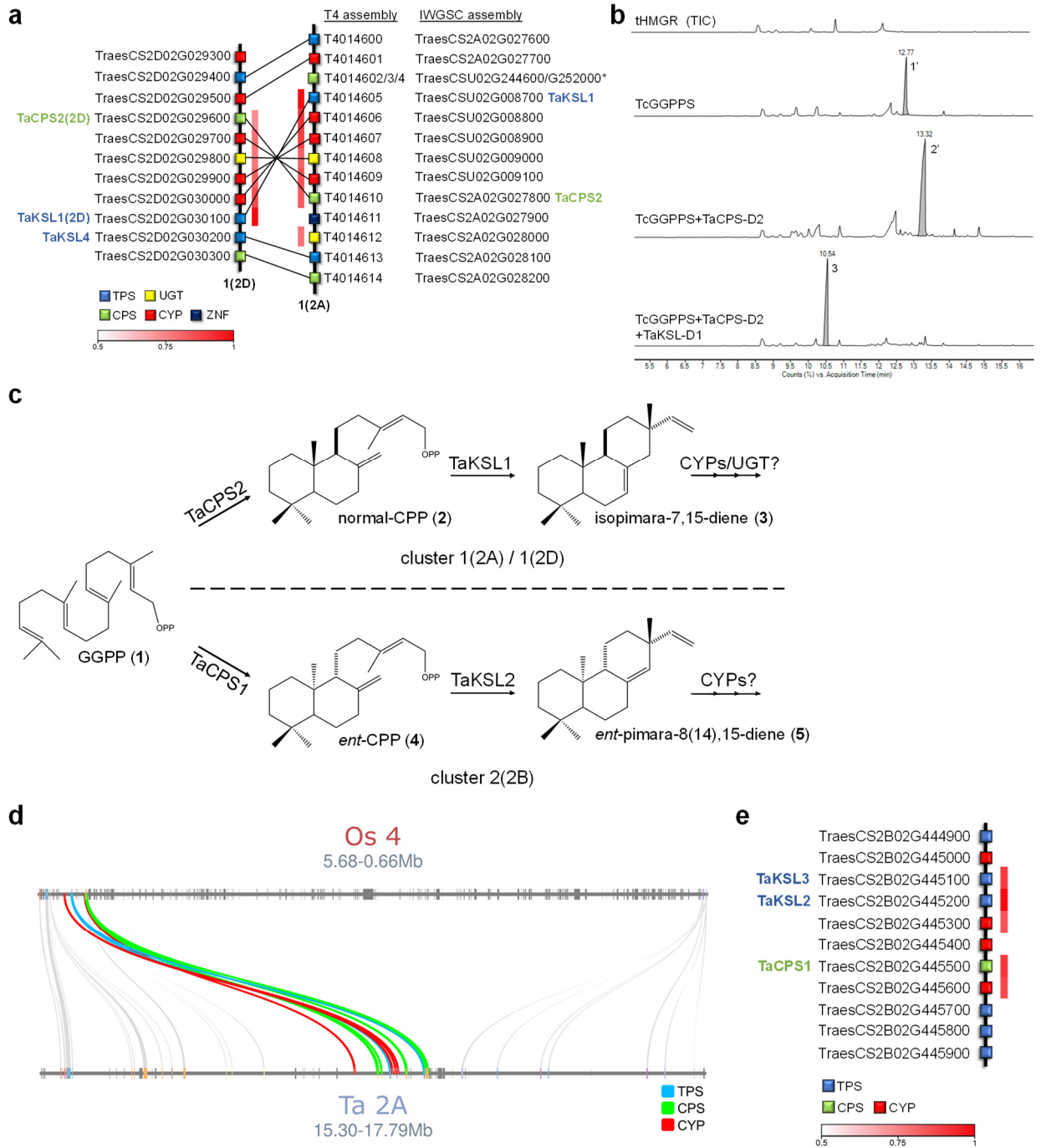
976 Figure 1



977

978 **Fig. 1: Co-expression network analysis reveals candidate defense-related BGCs in bread wheat.**
979 **a**, differential expression values of eigengenes representing 69 gene expression modules generated by
980 Weighted Gene Co-expression Network Analysis (WGCNA) of a ‘disease’ subset of wheat genes.
981 Eigengenes are sorted by mean of delta between treatment and control experiments with different
982 pathogens or PAMPs: Fpg1, *Fusarium pseudograminearum*⁷⁷; Fpg2, *Fusarium pseudograminearum*⁷⁸;
983 Ztr, *Zymoseptoria tritici*⁷⁹; Bgt, *Blumeria graminis f. sp. tritici*⁸⁰; Pst, *Puccinia striiformis f. sp. tritici*³⁹;
984 Chit, chitin⁸; Flg22, flagellin peptide⁸. **b**, normalized expression of eigengenes from modules ME34,
985 ME25, ME12, ME36 and ME8, and average of eigengenes from all modules of the ‘disease’ network.
986 The number of genes (n) within each module is indicated. Control and treatment experiments in each
987 study are represented by full and empty circles, respectively. **c**, putative disease-related biosynthetic
988 gene clusters (BGCs) identified in wheat. The red lines indicate the chromosomal positions of the BGCs
989 on the bread wheat chromosomes. The different types of cluster genes are colour-coded according to
990 their annotation: *TPS*, terpene synthase; *OSC*, oxidosqualene cyclase; *UGT*, UDP-glycosyltransferase;
991 *CYP*, cytochrome P450; *CPS*, copalyl diphosphate synthase; *OMT*, O-methyl transferase; *HSD*,
992 hydroxysteroid dehydrogenase; *CHI*, chalcone isomerase; *CHS*, chalcone synthase; *ZNF*, Zinc finger
993 RING/FYVE/PHD-type. Clusters are named according to type (types 1-4) and chromosomal location.
994 The white to red color-coding denotes the Pearson correlation (r) values for expression of each gene
995 with a representative ‘bait’ gene from the cluster. **d**, target gene-transcription factor interactions derived
996 from a GENIE3-based wheat regulatory network. Yellow nodes represent target genes from six BGCs.
997 Light blue nodes represent transcription factors interacting with one or more target genes. Pairwise
998 interaction weight is denoted by edge width. Representative TFs from the most highly interacting TF
999 groups (based on sum of all interaction weights) are enlarged and annotated: bHLH1
1000 (TraesCS3B01G122800), NAC (TraesCS5A01G411800), HSF (TraesCS1A01G350400), bHLH2
1001 (TraesCS7D01G360600), WRKY (TraesCS2D01G011700). **e**, LC-PDA analysis of the products of
1002 cluster 4(5D) genes following transient *Agrobacterium*-mediated expression in *N. benthamiana*. The
1003 product at Rt=20.3, marked with an arrow, is formed with expression of the complete cluster, but not
1004 in the absence of any of the genes *TaCYP71C164_5D*, *TaCYP71F53_5D*, *TaOMT3*, *TaOMT6*, or
1005 *TaOMT8*. pEAQ-HT-DEST1, empty vector control.

1006 **Figure 2**



1007

1008

1009

1010

1011

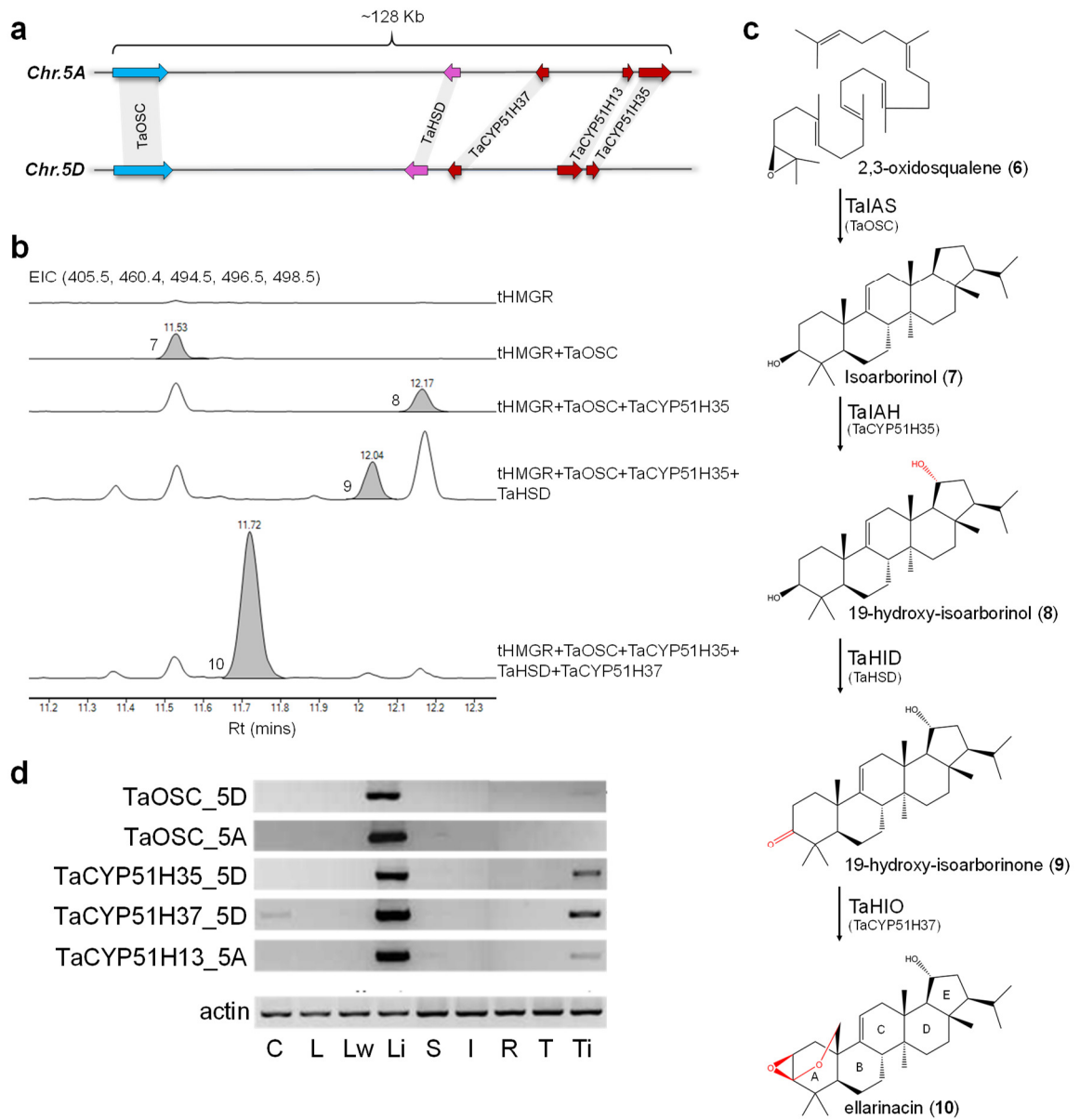
1012 **Fig. 2: Diterpene-producing BGCs are found on group 2 chromosomes in bread wheat. a,**
1013 assignment of homoeologous genes in the type 1 clusters 1(2A) and 1(2D), including the previously
1014 characterized genes *TaCPS2*, *TaKSL1* and *TaKSL4*. Chr.2A genes were positioned based on the T4
1015 wheat genome assembly⁵⁹ and homoeologs were assigned based on pairwise sequence alignments. The
1016 T4 assembly reveals the presence of five Chr.2D homoeologs in inverted positions on Chr.2A
1017 (*TraesCSU02G008700-G009100*), which were previously unmapped in the IWGSC assembly. *CPS*
1018 genes *TraesCSU02G252000* and *TraesCSU02G244600* (asterisked) have partial coding sequences.
1019 *TPS*, terpene synthase; *UGT*, UDP-dependent glycosyltransferase; *CPS*, copalyl diphosphate synthase;
1020 *CYP*, cytochrome P450; *ZNF*, Zinc finger, RING/FYVE/PHD-type. The white to red color-coding
1021 denotes Pearson correlation (r) values for expression of each gene with a representative ‘bait’ gene from
1022 the cluster. **b**, GC-MS analysis of leaf extracts following expression of the wheat TaKSL-D1 and
1023 TaCPS-D2 enzymes in *N. benthamiana*. Cytosol-targeted TaKSL-D1 and TaCPS-D2 were transiently
1024 expressed together with a *Taxus canadensis* GGPP synthase and oat tHMGR. Total ion chromatograms
1025 (TIC) are shown. Peaks were putatively identified as geranylgeraniol (**1'**), copalol (**2'**), and isopimara-
1026 7,15-diene (**3**), based on comparison of mass spectra to the NIST database and the literature (see
1027 Supplementary Fig. 3). **c**, predicted pathways for diterpene production by the type 1 BGCs 1(2A) and
1028 1(2D), and the type 2 BGC 2(2B). The type 1 clusters 1(2A) and 1(2D) comprise co-expressed genes
1029 for TaCPS2 and TaKSL1, CYPs and UGTs, predicted to form isopimara-7,15-diene-derived
1030 diterpenoids from geranylgeranyl diphosphate (GGPP). Cluster 2(2B), includes co-expressed genes for
1031 TaCPS1, TaKSL2, TaKSL3 and two CYPs, putatively forming pimara-8(14),15-diene-derived
1032 diterpenoids from GGPP. **d**, microsynteny analysis of wheat BGC 1(2A) (T4 assembly) and the
1033 momilactone cluster in a syntenic region in rice Chr.4. **e**, structure of BGC 2(2B) and assignment of the
1034 previously characterized genes *TaCPS1*, *TaKSL2* and *TaKSL3*.

1035

1036

1037

1038 **Figure 3**



1039

1040

1041

1042

1043

1044

1045 **Fig. 3: Wheat cluster 3(5D) produces an isoarborinol-derived triterpenoid. a**, structures of
1046 homoeologous triterpene biosynthetic gene clusters identified on wheat chromosomes 5A and 5D. **b**,
1047 GC-MS traces for wheat BGC 3(5D) genes transiently expressed in *N. benthamiana*. EIC, extracted ion
1048 chromatogram for ions representing isoarborinol (**7**) (498.5), 19-hydroxy-isoarborinol (**8**) (496.5), 19-
1049 hydroxy-isoarborinone (**9**) (494.5), ellarinacin (**10**) (405.5) and internal standard 5 α -cholestan-3 β -ol
1050 (460.4). **c**, assigned structure of ellarinacin and predicted biosynthetic pathway in wheat. TaIAS,
1051 isoarborinol synthase; TaIAH, isoarborinol 19-hydroxylase; TaHID, 19-hydroxy-isoarborinol
1052 dehydrogenase; TaHIO, 19-hydroxy-isoarborinone oxidase. Rings A-E are annotated. **d**,
1053 semiquantitative RT-PCR of selected genes from type 3 clusters 3(5A) and 3(5D) in ‘Chinese Spring’
1054 wheat tissues. C, coleoptile; L, leaf; L_w, leaf after wounding; L_i, leaf infected with *Blumeria graminis*
1055 *f. sp. tritici*; S, stem; I, inflorescence; R, root; T, root tip; T_i, root tip after infection with
1056 *Gaeumannomyces graminis*.

1057

1058

1059

1060

1061

1062

1063

1064

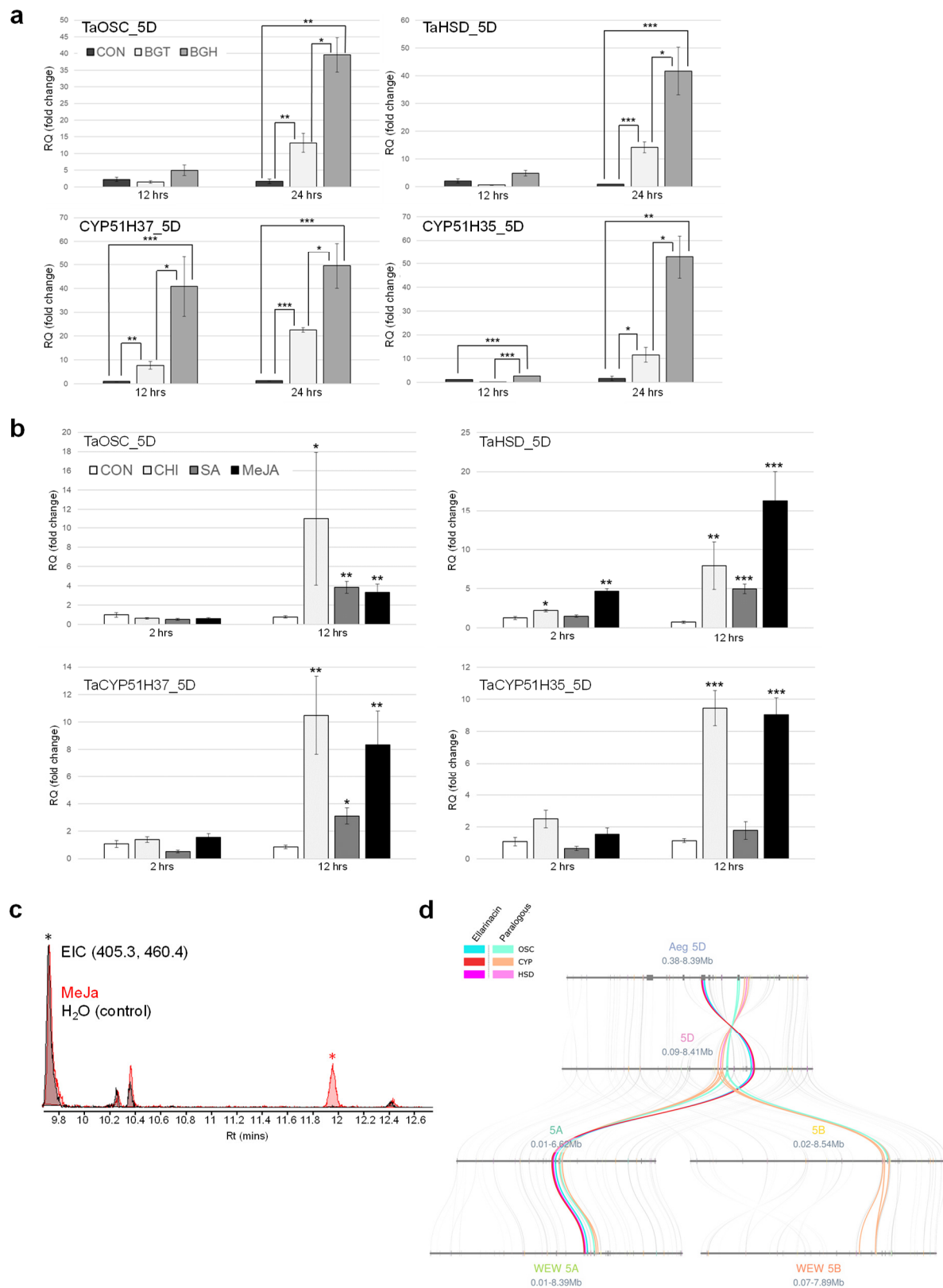
1065

1066

1067

1068

1069 **Figure 4**



1070

1071 **Fig. 4: The ellarinacin BGC 3(5D) is induced by pathogens and elicitors.** **a**, quantitative real-time
1072 PCR (qRT-PCR) of ellarinacin BGC genes in detached wheat leaves infected with two powdery mildew
1073 isolates, 12 and 24 hours post infection. Con, control (non-infected); Bgt and Bgh, infected with wheat-
1074 adapted isolate *Blumeria graminis f. sp. tritici* or the non-adapted isolate *Blumeria graminis f. sp.*
1075 *Hordei*, respectively **b**, quantitative real-time PCR (qRT-PCR) for ellarinacin BGC genes in detached
1076 wheat leaves treated with methyl jasmonate (MeJa), salicylic acid (SA), chitin (CHI) or H₂O (CON),
1077 for 2 or 12 hours. For panels A and B, relative quantification values (in fold-change) indicate means of
1078 three biological replicates \pm SEM. Asterisks denote t-test statistical significance of differential
1079 expression. *, p -val<0.05. **, p -val<0.01. ***, p -val<0.001. **c**, GC-MS analysis of TMS-derivatized
1080 extracts from wheat leaves treated with methyl jasmonate (MeJa), or H₂O (control) for three days.
1081 Extracted ion chromatograms are for ions representing ellarinacin (405.3, Rt 11.94, red asterisk) and
1082 5 α -cholestan-3 β -ol (460.4, Rt 9.70, black asterisk). **d**, microsynteny analysis of the region surrounding
1083 the ellarinacin BGC and its paralogous cluster in Chr.5 of the wheat A, B and D genomes, and wheat
1084 progenitors *Aegilops tauschii* (Aeg) and wild emmer wheat (WEW).

1085

1086

1087

1088

1089

1090

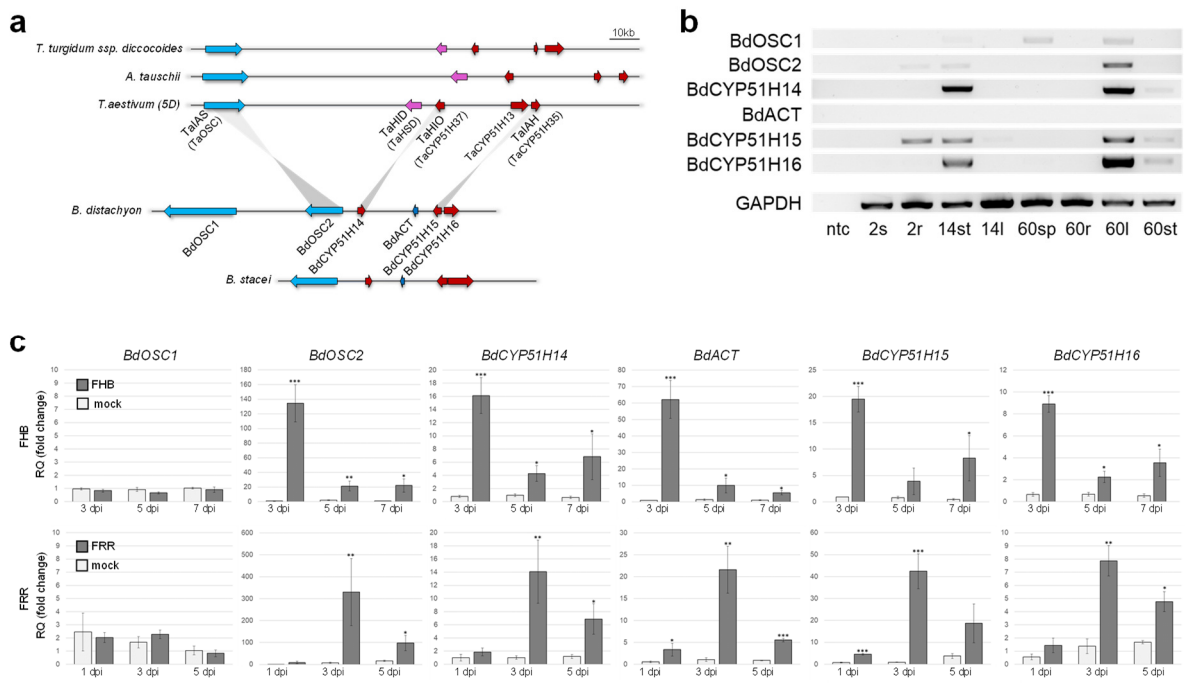
1091

1092

1093

1094

1095 **Figure 5**



1096

1097

1098 **Fig. 5: Occurrence and expression of ellarinacin-like BGCs in *Brachypodium* and wheat ancestral**

1099 **species. a**, wheat ellarinacin BGC is conserved in wheat wild ancestors *Aegilops tauschii* and wild

1100 emmer wheat (*Triticum turgidum subsp. dicoccoides*), and homologous to a BGC identified in

1101 chromosome 3 of *Brachypodium distachyon*. Grey lines link between wheat and *B. distachyon* BlastP

1102 reciprocal best hits. **b**, semiquantitative RT-PCR of *B. distachyon* Chr.3 clustered genes. ntc, no

1103 template control; 2s, seedling shoot (2 day old); 2r, seedling root; 14st, young plant stem base (14 day

1104 old); 14l, young plant leaf; 60sp, mature plant spike (60 day old); 60r, mature plant root; 60l, mature

1105 plant leaf; 60st, mature plant stem base. **c**, quantitative real-time PCR (qRT-PCR) of brachynacin BGC

1106 genes in *B. distachyon* plants infected with Fusarium head blight (FHB) or Fusarium root rot (FRR).

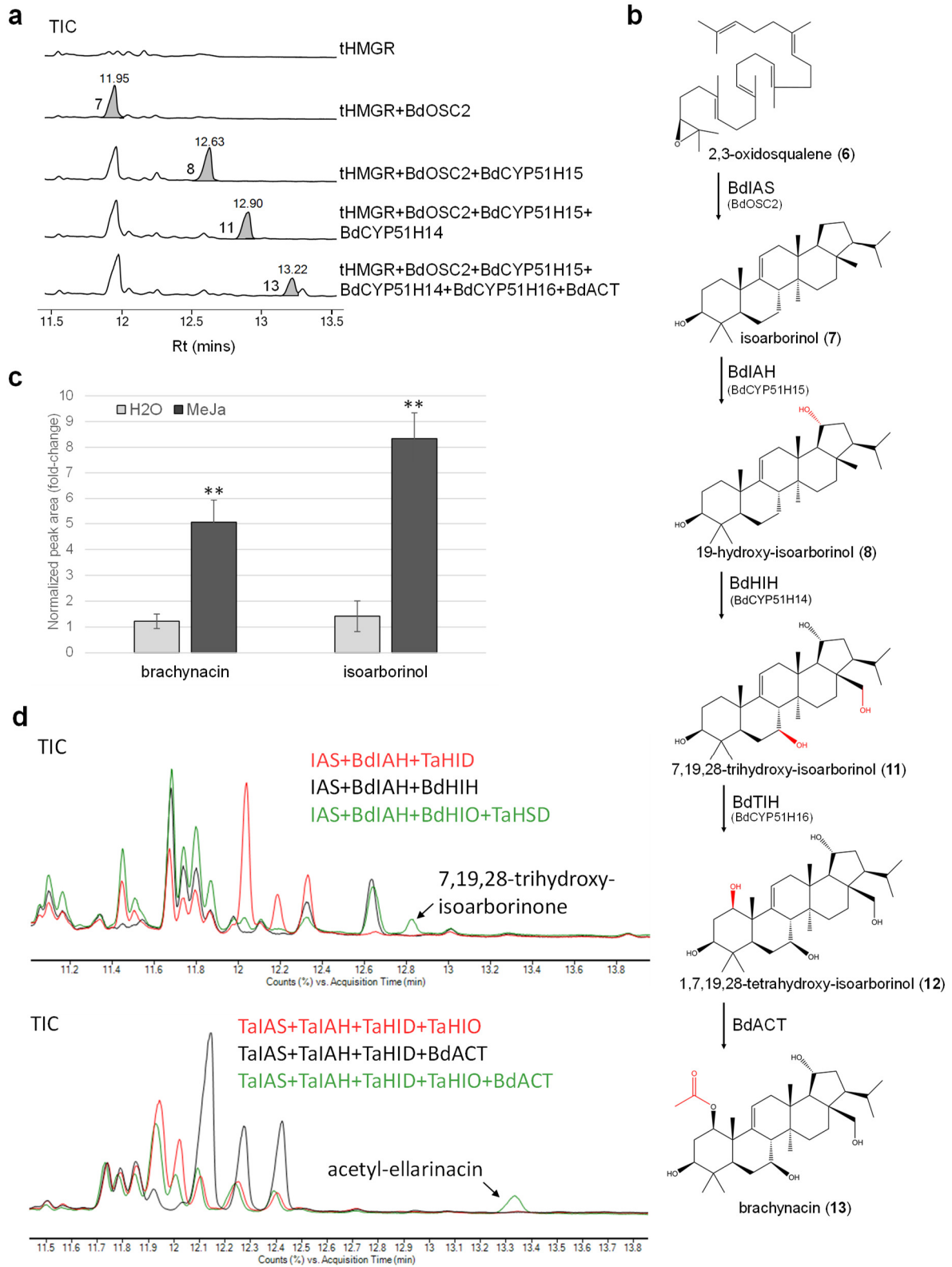
1107 Con, control (non-infected); dpi, days post infection. Relative quantification values (in fold-change)

1108 indicate means of three biological replicates \pm SEM. Asterisks denote t-test statistical significance of

1109 differential expression. *, p -val<0.05. **, p -val<0.01. ***, p -val<0.001.

1110

1111 **Figure 6**



1112

1113

1114 **Fig. 6: *B. distachyon* BGC produces the isoarborinol-derived triterpenoid, brachynacin. a,** GC-
1115 MS traces for *B. distachyon* cluster genes transiently expressed in *N. benthamiana*. TIC, total ion
1116 chromatogram. Marked peaks were identified as isoarborinol (**7**), 19-hydroxy-isoarborinol (**8**), 7,19,28-
1117 trihydroxy-isoarborinol (**11**) and brachynacin (**13**) (494.5). **b,** assigned structures and predicted
1118 biosynthetic pathway of brachynacin in *B. distachyon*. BdIAS, isoarborinol synthase; BdIAH,
1119 isoarborinol hydroxylase; BdHIH, 19-hydroxy-isoarborinol hydroxylase; BdTIH, 7,19,28-trihydroxy-
1120 isoarborinol hydroxylase. BdACT 1,7,19,28-tetrahydroxy-isoarborinol acetyltransferase. **c,** relative
1121 abundance of isoarborinol and brachynacin in TMS-derivatized extracts of *B. distachyon* leaves treated
1122 with MeJa or H₂O for 12 hours. Relative quantification is based on normalized peak areas in GC-MS
1123 analysis of four biological replicates. **d,** GC-MS total ion chromatograms (TIC) of *N. benthamiana*
1124 leaves transiently expressing combinations of wheat and *B. distachyon* genes.

1125

1126

1127

1128

1129

1130

1131

1132

1133

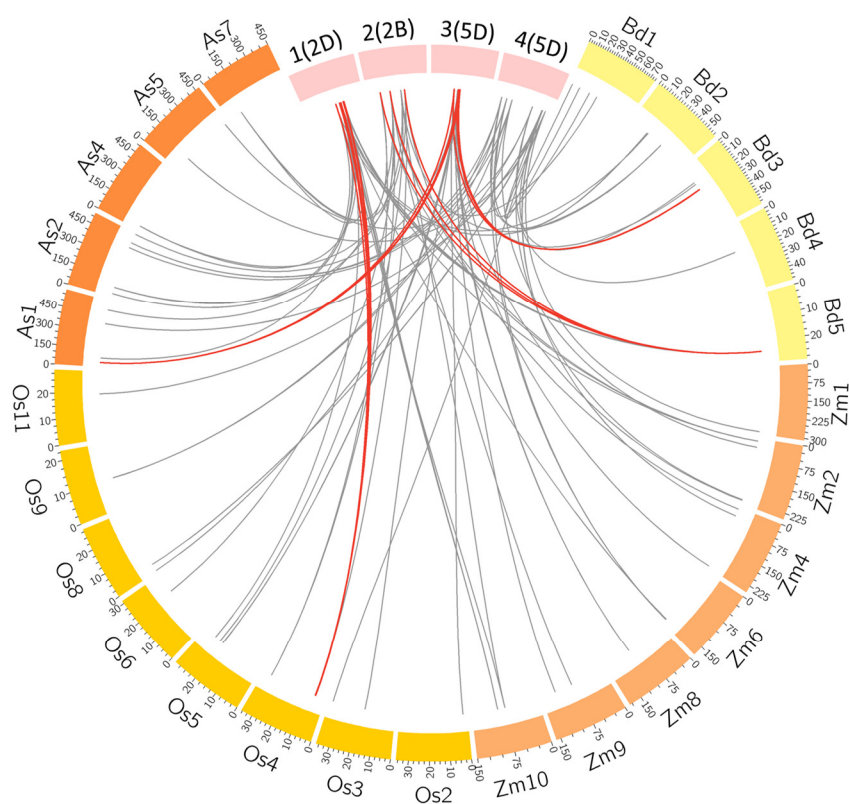
1134

1135

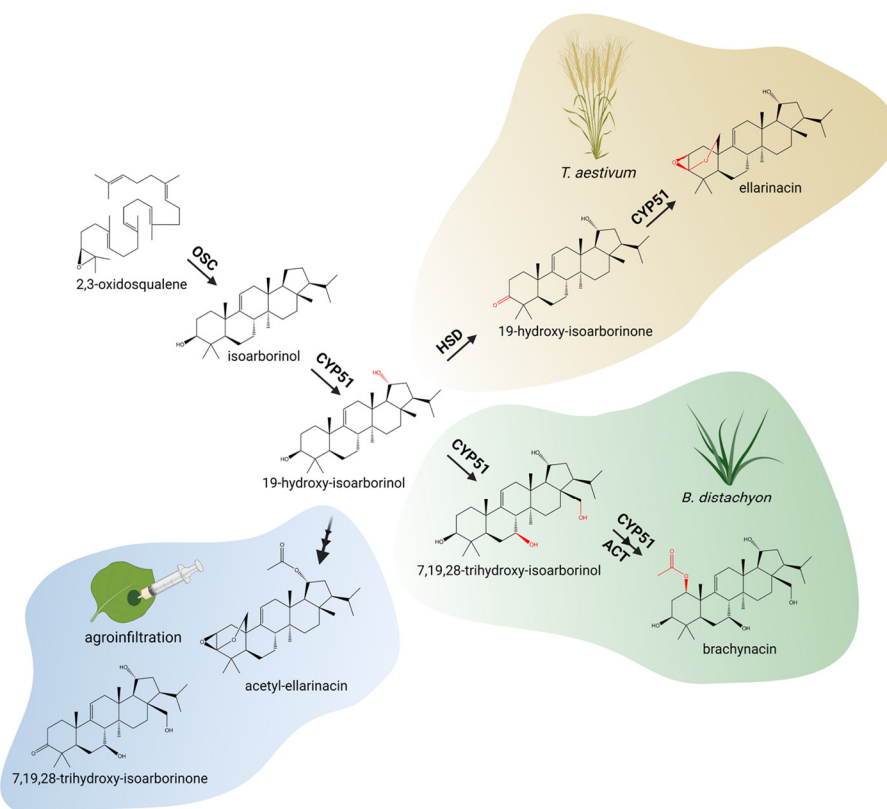
1136

1137 **Figure 7**

a



b



1138

1139 **Fig. 7: Phylogenetic and chemical divergence of arborinane-type biosynthetic gene clusters. a,**
1140 Circos plot depicting genomic locations of closest matching homologs of co-expressed genes from
1141 wheat BGCs 1(2D), 2(2B), 3(5D) and 4(5D) on chromosomes of *B. distachyon* (Bd), diploid oat *Avena*
1142 *strigosa* (As), maize (Zm) or rice (Os). In grey: links to homologs dispersed across the analyzed
1143 genomes. In red: links where two or more matching homologs from different gene families co-localize
1144 in the analyzed genomes. Cluster 1(2D) is linked to momilactone BGC on rice Chr.4; cluster 2(2B) is
1145 linked to a putative terpene BGC in *B. distachyon* Chr.5; cluster 3(5D) is linked to the brachynacin
1146 BGC in *B. distachyon* Chr.3 and OSC-CYP51H pair in oat Chr.1; cluster 4(5D) homologs are dispersed
1147 in all grass genomes included in the analysis. **b,** clustered biosynthetic pathways for arborinane-type
1148 triterpenoids in wheat and *B. distachyon* diverge from a common precursor, 19-hydroxy-isoarborinol,
1149 due to neofunctionalization of CYP51 enzymes and recruitment of other gene families. These pathways
1150 can be further artificially ‘diverged’ by recombinant expression of combined genes from the two
1151 clusters. Image created with Biorender.